(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064843 A1

(51) International Patent Classification[7]: C12R 1/00, C07D 241/36, C12N 15/00

(21) International Application Number: PCT/US02/03703

(22) International Filing Date: 8 February 2002 (08.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/267,663 9 February 2001 (09.02.2001) US
60/349,321 15 January 2002 (15.01.2002) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(71) Applicants and
(72) Inventors: **HAYGOOD, Margo** [US/US]; 2390 King Arthur Court, La Jolla, CA 92037 (US). **SALOMON, Christine** [US/US]; 9230F Regents Road, La Jolla, CA 92037 (US). **FAULKNER, John, D.** [US/US]; 5915 Desert View Drive, La Jolla, CA 92037 (US).

(74) Agent: **VIKSNINS, Ann, S.**; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, Minnesota 55402 (US).

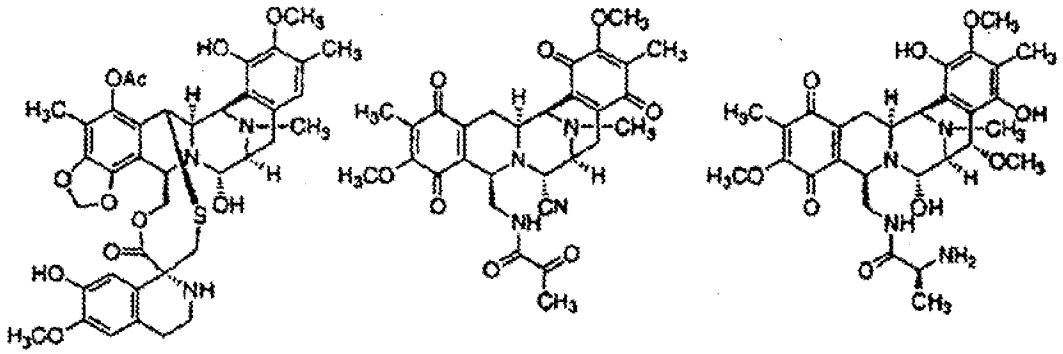
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ECTEINASCIDIN FAMILY COMPOUNDS: COMPOSITIONS AND METHODS



(57) Abstract: The present invention provides a bacterial symbiont of *Ecteinascidia turbinata*, and methods for the synthesis of an ecteinascidin.

WO 02/064843 A1

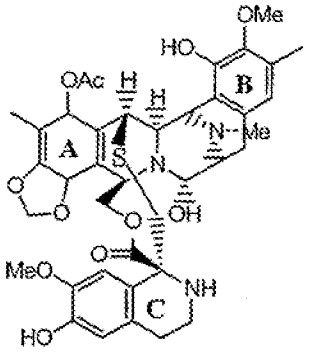
ECTEINASCIDIN FAMILY COMPOUNDS: COMPOSITIONS AND METHODS

Technical Field

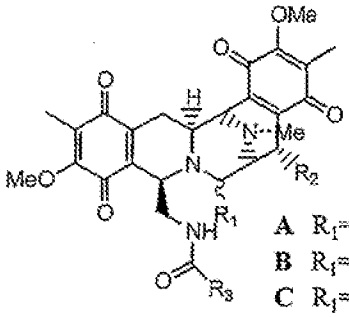
The present invention generally relates to the production of bioactive compounds such as ecteinascidins and methods of making such ecteinascidins using a variety of methods. In particular, bacterial symbionts of the ascidian *Ecteinascidia turbinata*, its proteins or genes can be used in such methods.

Background of the Invention

The ecteinascidins are a group of bioactive tetrahydroisoquinoline alkaloids isolated from the Caribbean mangrove ascidian *Ecteinascidia turbinata* (Rinehart et al., 1990 and Wright et al., 1990). The most active of these compounds, ecteinascidin 743 (Et-743), (Compound **1**) shows remarkable potency in both *in vitro* and *in vivo* bioassays, and possesses picomolar activity against breast, colon, non-small cell lung carcinoma (NSCLC), renal, ovarian, and melanoma cell lines (Zewail-Foote et al., 1999). The potent activity of Et-743 and its clinical potential places Et-743 in a small group of marine invertebrate-derived compounds, including discodermolide (Gunasekera et al., 1990), bryostatins (Pettit et al., 1982) and eleutherobin (Lindel et al., 1997).



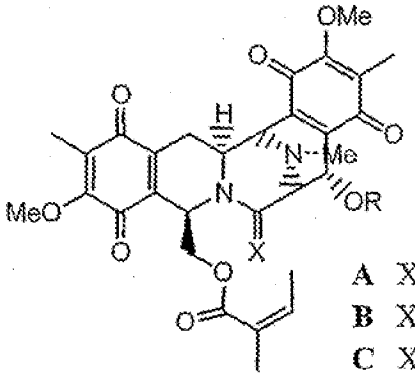
1 ecteinascidin 743 (Et 743)



A R_1=CN, R_2=H, R_3=$COCH_3$
B R_1=R_2=H, R_3=$COCH_3$
C R_1=H, R_2=OMe, R_3=$COCH_3$

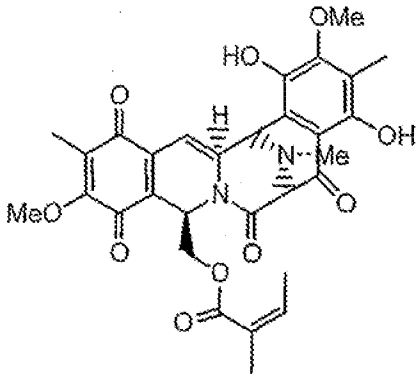
2 saframycins A-C

3 cyanosafracin R_1=CN, R_2=H, R_3=CNH_2CH_3

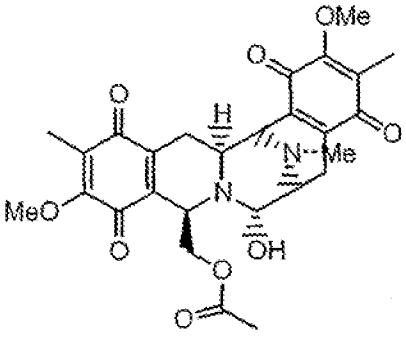


A X=H_2, R=H
B X=H_2, R=Et
C X=O, R=H
D X=O, R=Et

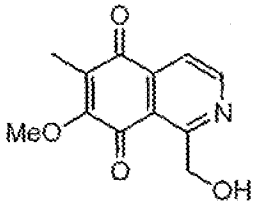
4 renieramycins A-D



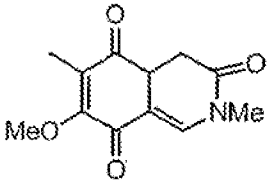
5 cribrostatin 4



6 jorunycin



7 renierol



8 mimosamycin

Et-743 binds to the minor groove of DNA and specifically alkylates the N2 of guanine (Pommier et al., 1996 and Moore et al., 1997). The drug differs from the other minor groove-interactive compounds because it induces a bend in the DNA helix toward the major groove (Zewail-Foote et al., 1999). Et-743 shows anti-tumor activity with IC_{50}s in the nanomolar range and is related to a large number of less potent compounds (Rinehart et al., 1990 and Faircloth et al., 1996). Et-743 is in phase I clinical trials in the United States and Europe and is in phase II clinical trials in France (Valori et al., 1988 and Rinehart, 2001).

Mechanistically, Et-743 disrupts the microtubule network and binds substantially irreversibly to 5'-GGG, 5'-GGC and 5'-AGC DNA sequences that leads to cell cycle arrest in the late S and G2/M phase (Garcia-Rocha et al., 1996;

Morre et al., 1997; Pommier et al., 1996; Morre et al., 1998; and Faircloth et al., 1996).

Structurally, Et-743 is similar to the renieramycins (He et al., 1989 and Frincke et al., 1982) from the marine sponge *Reniera* sp. It shows remarkable structural similarity to the less active saframycin antibiotics, (Compounds **2-3**), respectively, isolated from the bacteria *Streptomyces lavendulae* (Arai et al., 1983), *Pseudomonas fluorescens* (Meyers et al., 1983) and *Myxococcus xanthus* (Irschik et al., 1988), as well as the renieramycins (Compound 4) and cribrostatin 4 (Compound **5**) isolated from the marine sponges *Reniera* sp. and *Cribrochalina* sp., respectively (Frincke et al, 1982 and Pettit et al., 2000). Saframycin derivates are produced in culture and synthetic routes exist for derivatives of most of the isoquinolone antibiotics.

Jorumycin (Compound **6**) was isolated from the Indian nudibranch *Jorunna funebris* that was found "associated with" a blue sponge, identified as *Oceanapia* sp. (Fontana et al., 2000). The monomeric isoquinolines (Compounds 7 and **8**) were also found in the *Oceanapia* sponge, as well as culture broths of *S. lavendulae*, and a variety of other sponges (Frincke and Faulkner, 1982; McKee and Ireland, 1987; Parameswaran et al., 1998). It is believed that compounds such as Compounds 7 and 8 are the result of oxidative cleavage of the dimeric compounds and probably arise from the isolation and extraction procedure (Frincke and Faulkner, 1982). Et-743 differs from compounds **2-6** primarily by the presence of the C ring sub-unit system, which is critical for its high potency and tumor selectivity (Sakai et al., 1996).

Et-743 is expensive, since natural marine products are typically difficult to obtain in sufficient quantity for clinical use. Although aquaculture currently provides a source for Et-743, it yields only a small amount of the compound per mass of invertebrate (the average yield of the drug from the ascidian are generally low (0.01-0.02 % dry weight)). Until recently, the most efficient method of Et-743 production was aquaculture of *E. turbinata* (Corey et al., 1996). Efforts have been made to produce Et-743 more economically by synthetic processes (Martinez and Corey, 2000; Zhou et al., 2000; Herberich et

al., 2001). A relatively efficient chemical synthesis exists for the compound but is a less promising source than aquaculture (Corey et al., 1996). Cuevas et al. reported efficient semi-synthesis of Et-743 from cyanosafracin B (Compound 3) which can be easily obtained in kilogram quantities from culture of *Pseudomonas fluorescens* (Cuevas et al., 2000). An attractive alternative to aquaculture or even semi-synthesis may be the cloning of the biosynthetic genes into a more tractable host, such as *P. fluorescens*. Thus, there is a need for alternative methods for producing ecteinascidins, such as Et-743.

Brief Description of the Figures

FIG. 1 depicts the structure of Ecteinascidin-743 (Et-743).

FIG. 2 depicts the modification of L-Tyr.

FIG. 3 depicts a scheme for the synthesis of Et-743.

FIG. 4 depicts *in situ* hybridization of bacterial 16S rRNA probes to larvae of *E. turbinata*.

FIG. 5 depicts adult *E. turbinata* showing brooding of eggs in lower portion of the zooid. (15X) (a); excised larvae at various development stages (20X) (b).

FIG. 6 depicts (a) ESI mass spectrum from the *E. turbinata* larval extract; (b) MS/MS of the parent ion (744) showing loss of the C-ring subunit of Et-743.

FIG. 7 depicts DGGE gel of PCR products amplified with 1055f-1392r. 1: *E. turbinata*, 2. *E. turbinata*, 3. *E. coli*. Arrows point to the two bands that were excised and sequenced.

FIG. 8 depicts phylogram obtained with maximum parsimony and 100x bootstrapping (indicated by numbers).

FIG. 9 depicts FISH experiments using fluorescein labeled probes. A. Domain Bacteria probe Eub338; B. no probe control; C. *E. palauensis* probe 1247r; D. *E. sertula* probe 1253r.

FIG. 10 depicts FISH micrographs of *E. turbinata* larvae using all excitation wavelengths on a confocal microscope. A. Whole larva hybridized with CY5 labeled Eub338 (indicated by blue) 10X, scale bar = 100 μm. (B-D)

Images of larvae taken near adhesive papillae B. 40X, scale bar = 20 μm. C. 63X, scale bar = 10 μm. D. 63X plus digital zoom, scale bar = 5 μm.

FIG. 11 depicts PCR amplification of *E. turbinata* larval DNA using a combination of universal, *Endobugula* specific and *E. sertula* specific primers. 1 = *E. turbinata larval* DNA, 2 = *B. pacifica* DNA used as a control, 3 = negative control. A: universal bacterial primers (27f-1492r), B: universal with *Endobugula* specific (27f-855r), C: *E. sertula* specific (240f-1253r), D: universal with *E. sertula* specific (27f-1253r).

FIG. 12 depicts light micrograph sections of *E. turbinata* larva through the adhesive papillae. A. 16X, scale bar = 70 μm B. 40X, scale bar = 30 μm.

Summary of the Invention

Methods provided herein include using bacteria that are involved at least in part in the synthesis of bioactive compounds. These bacteria can be used in production processes, such as fermentation, to produce valuable bioactive compounds. Alternatively, polypeptides present in such bacterial or nucleic acid molecules that encode such polypeptides can be used to produce bioactive compounds in *ex vivo* processes or in recombinant organisms engineered to express such polypeptides. The present invention provides these and related benefits.

The present invention recognizes that organisms such as bacteria contain nucleic acid molecules that encode polypeptides that catalyze the synthesis of bioactive compounds, such as ecteinascidins.

Bacteria

One aspect of the present invention is a bacterium or bacteria that can be a pure or mixed culture, that at least in part synthesizes at least one ecteinascidin.

A second aspect of the present invention is a preparation of bacteria from an *Ecteinascidia*, such as *Ecteinascidia turbinata*, that includes at least one activity involved in the synthesis of an ecteinascidin.

Nucleic Acids

A third aspect of the present invention is a composition including at least one nucleic acid molecule that encodes at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin.

A fourth aspect of the present invention is a composition including a library of nucleic acid molecules of the present invention. These nucleic acid molecules can be used in a combinatorial biosynthesis of ecteinascidins.

Polypeptides

A fifth aspect of the present invention is a composition including at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin, where preferably at least one polypeptide is derived from at least one marine organism.

A sixth aspect of the present invention is a composition including a library of polypeptides of the present invention.

Antibodies

A seventh aspect of the present invention is an antibody that binds with or specifically binds with one or more bacteria of the present invention.

An eighth aspect of the present invention is an antibody that binds with or specifically binds with a polypeptide of the present invention.

Method of making bioactive compounds

Bacteria

A ninth aspect of the present invention is a method of making an ecteinascidin using a bacteria of the present invention.

A tenth aspect of the present invention is a composition made using a bacteria of the present invention.

Nucleic Acids

An eleventh aspect of the present invention is a method of making a composition, including providing at least one nucleic acid of the present invention and synthesizing at least one ecteinascidin.

A twelfth aspect of the present invention is a composition made using a nucleic acid of the present invention.

Polypeptides

A thirteenth aspect of the present invention is a method of making a composition including providing at least one polypeptide of the present invention and synthesizing at least one ecteinascidin.

A fourteenth aspect of the present invention is a composition made using a polypeptide of the present invention.

Method for identifying nucleic acid molecules

A fifteenth aspect of the present invention is a method for identifying at least one nucleic acid molecule encoding at least one activity involved in the synthesis of an ecteinascidin, comprising contacting a nucleic acid molecule of the present invention with a sample, and identifying nucleic acid molecules in said sample that hybridize with a nucleic acid of the present invention.

A sixteenth aspect of the present invention is a nucleic acid molecule identified by a method of the present invention.

A seventeenth aspect of the present invention is a composition comprising a library of nucleic acid molecules identified using a method of the present invention.

Method for identifying bioactive compounds

Bacteria

An eighteenth aspect of the present invention is a method for identifying a bioactive compound including contacting a compound made or identified using a bacteria of the present invention with a detection system and determining the bioactivity of the compound.

A nineteenth aspect of the present invention is a compound identified by this method.

Nucleic Acid Molecule

A twentieth aspect of the present invention is a method for identifying a bioactive compound including contacting a compound made or identified using a nucleic acid molecule of the present invention with a detection system and determining the bioactivity of the compound.

A twenty-first aspect of the present invention is a compound identified by this method.

Polypeptide

A twenty-second aspect of the present invention is a method for identifying a bioactive compound including contacting a compound made or identified using a polypeptide of the present invention with a detection system and determining the bioactivity of said compound.

A twenty-third aspect of the present invention is a compound identified by this method.

Detailed Description of the Invention

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein and the laboratory procedures in cell culture, chemistry, microbiology, molecular biology, cell science and cell culture described below are well known and commonly employed in the art. Conventional methods are used for these procedures, such as those provided in the art and various general references (Sambrook et al., 1989; Ausubel et al., 1998; Harlowe and Lane, 1988). Where a term is provided in the singular, the inventors also contemplate the plural of that term. The nomenclature used herein and the laboratory procedures described below are those well known and commonly employed in the art. As employed throughout the disclosure, the following terms, unless otherwise indicated, shall be understood to have the following meanings:

"Membrane permeant derivative" refers to a chemical derivative of a compound that increases membrane permeability of the compound. These derivatives are made better able to cross cell membranes because hydrophilic groups are masked to provide more hydrophobic derivatives. Also, the making groups can be designed to be cleaved from the compound within a cell to make the compound more hydrophilic once within the cell. Because the substrate is

more hydrophilic than the membrane permeant derivative, it preferentially localizes within the cell (U.S. Patent No. 5,741,657 to Tsien et al., issued April 21, 1998).

"Isolated polynucleotide" refers to a polynucleotide of genomic, cDNA, PCR or synthetic origin, or some combination thereof, which by virtue of its origin, the isolated polynucleotide (1) is not associated with the cell in which the isolated polynucleotide is found in nature, or (2) is operably linked to a polynucleotide that it is not linked to in nature. The isolated polynucleotide can optionally be linked to promoters, enhancers, or other regulatory sequences.

"Isolated protein" refers to a protein of cDNA, recombinant RNA, or synthetic origin, or some combination thereof, which by virtue of its origin the isolated protein (1) is not associated with proteins normally found within nature, or (2) is isolated from the cell in which it normally occurs, or (3) is isolated free of other proteins from the same cellular source, for example, free of cellular proteins), or (4) is expressed by a cell from a different species, or (5) does not occur in nature.

"Polypeptide" is used herein as a generic term to refer to native protein, fragments, or analogs of a polypeptide sequence.

"Active fragment" refers to a fragment of a parent molecule, such as an organic molecule, nucleic acid molecule, or protein or polypeptide, or combinations thereof, that retains at least one activity of the parent molecule.

"Naturally occurring" refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism, including viruses, that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally occurring.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence operably linked to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

"Control sequences" refer to polynucleotide sequences that effect the expression of coding and non-coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal biding site, and transcription termination sequences; in eukaryotes, generally, such control sequences include promoters and transcription termination sequences. The term control sequences is intended to include components whose presence can influence expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

"Polynucleotide" refers to a polymeric form of nucleotides of a least ten bases in length, either ribonucleotides or deoxyribonucleotides or a modified from of either type of nucleotide. The term includes single and double stranded forms of DNA or RNA.

"Genomic polynucleotide" refers to a portion of the genome.

"Active genomic polynucleotide" or active portion of a genome" refer to regions of a genome that can be up regulated, down regulated or both, either directly or indirectly, by a biological process.

"Directly" in the context of a biological process or processes, refers to direct causation of a process that does not require intermediate steps, usually caused by one molecule contacting or binding to another molecule (the same type or different type of molecule). For example, molecule A contacts molecule B, which causes molecule B to exert effect X that is part of a biological process.

"Indirectly" in the context of a biological process or precesses, refers to indirect causation that requires intermediate steps, usually caused by two or more direct steps. For example, molecule A contacts molecule B to exert effect X which in turn causes effect Y.

"Sequence identity" refers to the proportion of base matches between two nucleic acid sequences or the proportion of amino acid matches between two amino acid sequences. When sequence identity is expressed as a percentage, for example 50%, the percentage denotes the proportion of matches of the length of

sequences from a desired sequence that is compared to some other sequence. Gaps (in either of the two sequences) are permitted to maximize matching; gap lengths of 15 bases or less are usually used, 6 bases or less are preferred with 2 bases or less more preferred. When using oligonuleotides as probes, the sequence identity between the target nucleic acid and the oligonucleotide sequence is preferably not less than 10 target base matches out of 20 (50% identity) and more preferably not less than about 60% identity, 70% identity, 80% identity or 90% identity), and most preferably not less than 95% identity.

"Selectively hybridize" refers to detectably and specifically bind. Polynucleotides, oligonucleotides and fragments thereof selectively hybridize to target nucleic acid strands, under hybridization and wash conditions that minimize appreciable amounts of detectable binding to nonspecific nucleic acids. High stringency conditions can be used to achieve selective hybridization conditions as known in the art. Generally, the nucleic acid sequence identity between the polynucleotides, oligonucleotides, and fragments thereof and a nucleic acid sequence of interest will be at least 30%, and more typically and preferably of at least 40%, 50%, 60%, 70%, 80% or 90%.

Hybridization and washing conditions are typically performed at high stringency according to conventional hybridization procedures. Positive clones are isolated and sequenced. For example, a full length polynucleotide sequence can be labeled and used as a hybridization probe to isolate genomic clones from an appropriate target library as they are known in the art. Typical hybridization conditions and methods for screening plaque lifts and other purposes are known in the art (Benton and Davis, Science 196:180 (1978); Sambrook et al., supra, (1989)).

Two amino acid sequences have share identity if there is a partial or complete identity between their sequences. For example, 85% identity means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred. Alternatively and preferably, two protein sequences

(or polypeptide sequences derived from them of at least 30 amino acids in length) share identity, as this term is used herein, if they have an alignment score of at least 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater (Dayhoff, in Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, volume 5, pp. 101-110 (1972) and Supplement 2, pp. 1-10).

"Corresponds to" refers to a polynucleotide sequence that shares identity (for example is identical) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to all or a portion of a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence TATAC corresponds to a reference sequence TATAC and is complementary to a reference sequence GTATA.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence," "comparison window," "sequence identity," "percentage of sequence identity," and "substantial identity." A reference sequence is a defined sequence used as a basis for a sequence comparison; a reference sequence can be a subset of a larger sequence, for example, as a segment of a full length cDNA or gene sequence given in a sequence listing, or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides can each (1) comprise a sequence (for example a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A comparison window, as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions

wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window can comprise additions and deletions (for example, gaps) of 20 percent or less as compared to the reference sequence (which would not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window can be conducted by the local identity algorithm (Smith and Waterman, Adv. Appl. Math., 2:482 (1981)), by the identity alignment algorithm (Needleman and Wunsch, J. Mol. Bio., 48:443 (1970)), by the search for similarity method (Pearson and Lipman, Proc. Natl. Acid. Sci. U.S.A. 85:2444 (1988)), by the computerized implementations of these algorithms such as GAP, BESTFIT, FASTA and TFASTA (Wisconsin Genetics Software Page Release 7.0, Genetics Computer Group, Madison, WI), or by inspection. Preferably, the best alignment (for example, the result having the highest percentage of identity over the comparison window) generated by the various methods is selected.

"Complete sequence identity" means that two polynucleotide sequences are identical (for example, on a nucleotide-by-nucleotide basis) over the window of comparison.

"Percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (for example, the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

"Substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 30 percent sequence identity, preferably at least 50 to 60 percent sequence, more usually at least 60 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25 to 50 nucleotides, wherein the

percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence that may include deletions or addition which total 20 percent or less of the reference sequence over the window of comparison.

"Substantial identity" as applied to polypeptides herein means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 30 percent sequence identity, preferably at least 40 percent sequence identity, and more preferably at least 50 percent sequence identity, and most preferably at lest 60 percent sequence identity. Preferably, residue positions, which are not identical, differ by conservative amino acid substitutions.

"Conservative amino acid substitutions" refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine and tryptophan; a group of amino acids having basic side chains is lysine, arginine and histidine; and a group of amino acids having sulfur-containing side chan is cystein and methionine. Preferred conservative amino acid substitution groups are: valine-leucine-isoleucine; phenylalanine-tyrosine; lysine-arginine; alanine-valine; glutamic-aspartic; and asparagine-glutamine.

"Modulation" refers to the capacity to either enhance or inhibit a functional property of a biological activity or process, for example, enzyme activity or receptor binding. Such enhancement or inhibition may be contingent on the occurrence of a specific event, such as activation of a signal transduction pathway and/or may be manifest only in particular cell types.

"Modulator" refers to a chemical (naturally occurring or non-naturally occurring), such as a biological macromolecule (for example, nucleic acid, protein, non-peptide or organic molecule) or an extract made from biological materials, such as prokaryotes, bacteria, eukaryotes, plants, fungi, multicellular

organisms or animals, invertebrates, vertebrates, mammals and humans, including, where appropriate, extracts of: whole organisms or portions of organisms, cells, organs, tissues, fluids, whole cultures or portions of cultures, or environmental samples or portions thereof. Modulators are typically evaluated for potential activity as inhibitors or activators (directly or indirectly) of a biological process or processes (for example, agonist, partial antagonist, partial agonist, antagonist, antineoplastic, cytotoxic, inhibitors of neoplastic transformation or cell proliferation, cell proliferation promoting agents, antiviral agents, antimicrobial agents, antibacterial agents, antibiotics, and the like) by inclusion in assays described herein. The activity of a modulator may be known, unknown or partially known.

"Test chemical" refers to a chemical or extract to be tested by at least one method of the present invention to be a putative modulator. A test chemical is usually not known to bind to the target of interest. "Control test chemical" refers to a chemical known to bind to the target (for example, a known agonist, antagonist, partial agonist or inverse agonist). Test chemical does not typically include a chemical added to a mixture as a control condition that alters the function of the target to determine signal specificity in an assay. Such control chemicals or conditions include chemicals that (1) non-specifically or substantially disrupt protein structure (for example denaturing agents such as urea or guandium, sulfhydryl reagents such as dithiotritol and beta-mercaptoethanol), (2) generally inhibit cell metabolism (for example mitochondrial uncouples) and (3) non-specifically disrupt electrostatic or hydrophobic interactions of a protein (for example, high salt concentrations or detergents at concentrations sufficient to non-specifically disrupt hydrophobic or electrostatic interactions). The term test chemical also does not typically include chemicals known to be unsuitable for a therapeutic use for a particular indication due to toxicity of the subject. Usually, various predetermined concentrations of test chemicals are used for determining their activity. If the molecular weight of a test chemical is known, the following ranges of concentrations can be used: between about 0.001 micromolar and about 10 millimolar, preferably between

about 0.01 micromolar and about 1 millimolar, more preferably between about 0.1 micromolar and about 100 micromolar. When extracts are uses a test chemicals, the concentration of test chemical used can be expressed on a weight to volume basis. Under these circumstances, the following ranges of concentrations can be used: between about 0.001 micrograms/ml and about 100 milligram/ml, preferably between about 0.01 micrograms/ml and about 10 milligrams/ml, and more preferably between about 0.1 micrograms/ml and about 1 milligrams/ml or between about 1 microgram/ml and about 100 micrograms/ml.

"Target" refers to a biochemical entity involved in a biological process. Targets are typically proteins that play a useful role in the physiology or biology of an organism. A therapeutic chemical typically binds to a target to alter or modulate its function. As used herein, targets can include, but not be limited to, cell surface receptors, G-proteins, G-protein coupled receptors, kinases, phosphatases, ion channels, lipases, phosholipases, nuclear receptors, intracellular structures, tubules, tubulin, and the like.

"Label" or "labeled" refers to incorporation of a detectable marker, for example by incorporation of a radiolabled compound or attachment to a polypeptide of moieties such as biotin that can be detected by the binding of a section moiety, such as marked avidin. Various methods of labeling polypeptide, nucleic acids, carbohydrates, and other biological or organic molecules are known in the art. Such labels can have a variety of readouts, such as radioactivity, fluorescence, color, chemiluminescence or other readouts known in the art or later developed. The readouts can be based on enzymatic activity, such as beta-galactosidase, beta-lactamase, horseradish peroxidase, alkaline phosphatase, luciferase; radioisotopes such as ^{3}H, ^{14}C, ^{35}S, ^{125}I or ^{131}I); fluorescent proteins, such as green fluorescent proteins; or other fluorescent labels, such as FITC, rhodamine, and lanthanides. Where appropriate, these labels can be the product of the expression of reporter genes, as that term is understood in the art. Examples of reporter genes are beta-lactamase (U.S. Patent No. 5,741,657 to Tsien et al., issued April 21, 1998) and green fluorescent

protein (U.S. Patent No. 5,777,079 to Tsien et al., issued July 7, 1998; U.S. Patent No. 5,804,387 to Cormack et al., issued September 8, 1998).

"Substantially pure" refers to an object species or activity that is the predominant species or activity present (for example on a molar basis it is more abundant than any other individual species or activities in the composition) and preferably a substantially purified fraction is a composition wherein the object species or activity comprises at least about 50 percent (on a molar, weight or activity basis) of all macromolecules or activities present. Generally , as substantially pure composition will comprise more than about 80 percent of all macromolecular species or activities present in a composition, more preferably more than about 85%, 90%, 95% and 99%. Most preferably, the object species or activity is purified to essential homogeneity, wherein contaminant species or activities cannot be detected by conventional detection methods) wherein the composition consists essentially of a single macromolecular species or activity. The inventors recognize that an activity may be caused, directly or indirectly, by a single species or a plurality of species within a composition, particularly with extracts.

A "ecteinascidin synthesis activity" is at least one activity involved in the synthesis of an ecteinascidin. Examples of ecteinascidin synthesis activity can include without limitation a non-ribosomal peptide synthetase activity, an O-methyl transferase activity, or a transaminase activity.

"Pharmaceutical agent or drug" refers to a chemical, composition or activity capable of inducing a desired therapeutic effect when property administered by an appropriate dose, regime, route of administration, time and delivery modality.

"Pharmaceutical agent or drug" refers to a chemical, composition or activity capable of inducing a desired therapeutic effect when property administered by an appropriate dose, regime, route of administration, time and delivery modality.

A "bioactive compound" refers to a compound that exhibits at least one bioactivity. A bioactive compound may be an organic molecule or inorganic

molecule, protein, lipid, carbohydrate, or nucleic acid, or any combination thereof, including complexes of any of these compounds.

A "bioactivity" refers to at least one activity that modulates a biological process. Preferred bioactivities include, but are not limited to: antitumor activity, anticancer cell activity, cytotoxic activity, antibacterial activity, antimicrobial activity, not being substantially susceptible to multi-drug resistance, antiviral activity, immunomodulatory activity, anti-inflammatory activity, and radiation protective activity.

"An ecteinascidin" refers to a compound that includes a tetrahydroisoquinoline alkaloid derived from the marine tunicate *Ecteinascidia turbinata* and has at least one bioactivity.

"Made at least in part" or "at least in part synthesized" means that an one or more entities participate in one or more steps in the synthesis of the compound of interest or precursor of a compound of interest but may not be the sole entity or entities that participate in the synthesis of the compound or precursor of interest. When a single step in the synthesis of a compound of interest or precursor of a compound of interest is performed by one or more entities, "made at least in part" can refer to the participation of one or more entities in a synthetic steps or process that also requires the participation of at least one other entity.

A "bioactive derivative" refers to a modification of a bioactive compound or bioactivity that retains at least one characteristic activity of the parent compound.

A "bioactive precursor" refers to a precursor of a bioactive compound or bioactivity that exhibits at least one characteristic activity of the resulting bioactive compound or bioactivity.

An "antibacterial activity" refers to an activity that reduces the growth rate or numbers of living in a sample, such as a culture of bacteria or a sample that includes at least one bacteria, including a patient. Such antibacterial activity can be directed against Gram-negative or Gram-positive bacteria and can be screened for or confirmed using methods known in the art.

An "antimicrobial activity" refers to an activity that reduces the growth rate or numbers of living microbes in a sample (including prokaryotic and/or eukaryotic microbes), such as a culture of microbes or a sample that includes at least one microbe, including a patient, and can be screened for or confirmed using methods known in the art.

An "antiviral activity" refers to an activity that reduces the infectivity of virus particles in a sample, such as in a sample including at least one virus, including a patient. Such antiviral activity can be directed against, from example, DNA or RNA containing viruses, including, but not limited to, herpesvirus, hepatitis virus and retrovirus. Such activity can be screened for using methods known in the art.

An "antitumor activity" refers to an activity that reduces the growth rate or number of tumor cells in a sample, such as a culture of tumor cells or a sample that includes at least one tumor cell, including a patient. Such antitumor activity can be directed against any type of tumor or tumor cell, including, but not limited to renal tumor, lung tumor, colon tumor, central nervous system tumor, melanoma, ovarian tumor and breast tumor.

An "anticancer cell activity" refers to an activity that reduces the growth rate or number of cancer cells in a sample, such as a culture of cancer cells or a sample that includes at least one cancer cell, including a patient. Such anticancer cell activity can be directed against any type of cancer cell, including, but not limited to renal cancer, leukemia, lung cancer, colon cancer, central nervous system cancer, melanoma, ovarian cancer and breast cancer.

A "cytotoxic activity" refers to an activity that reduces the number of viable cells in a sample, including prokaryotic cells, eukaryotic cells or both. Such activity can be screened using methods known in the art.

A "patient" or "subject" refers a whole organism in need of treatment, such as a farm animal, companion animal or human. An animal refers to any non-human animal.

Other technical terms used herein have their ordinary meaning in the art that they are used, as exemplified by a variety of technical dictionaries, such as

the McGraw-Hill Dictionary of Chemical Terms and the Stedman's Medical Dictionary.

The biological and ecological functions of the ecteinascidins are unknown. Some of the roles generally ascribed to secondary metabolites include deterrence against predation, inhibition of the settlement of fouling organisms, protection from bacterial or viral infections, and prevention of encroachment by adjacent organisms.

Localization studies of compounds in sponges or ascidians have depended upon the presence of unique chemical structural features, such as heavy elements or pH dependent fluorescence, high bioactivity, and/or the ability to separate morphologically distinct bacterial and eukaryotic cells (Thompson et al., 1983; Bewley et al., 1996; Salomon et al., 2001; and Unson et al., 1994). However, many natural products do not have distinctive structural characteristics, and may not have high activity in bioassays. In addition, some invertebrates are not amenable to tissue dissociation and cell enrichment, and many bacteria do not have unique morphologies. Invertebrates containing interesting natural products are particularly difficult to work with because many of them cannot be cultured easily in a laboratory and often must be collected and preserved in remote, tropical areas. Additionally, sponge and ascidian systems are complicated by the fact that the adults filter seawater for food and generally contain large quantities of dietary-derived bacteria within their mesophyl matrix and gut, respectively. In order to study the potential role of symbionts in the biosynthesis of specific compounds in such complicated systems, it is necessary to employ a combination of chemical, biological and genetic techniques.

One method of studying symbionts in typically complex invertebrate/microbial associations is to focus on the chemistry and bacterial populations of non-feeding larvae. Many invertebrate larvae swim for a short period of time (minutes to days) and often do not filter feed until they permanently settle onto a surface and metamorphose into an adult (Strathmann, 1985). Many invertebrates associated with obligate symbionts are known to transmit bacteria vertically to their offspring; therefore, any bacteria that are

consistently and specifically associated with the larvae are presumed to be true symbionts (Krueger et al., 1996; Benayahu and Schleyer, 1998; Hirose, 2000). This approach was used to implicate a bacterial symbiont as a source of the potential anti-cancer agents bryostatins from the bryozoan *Bugula neritina*. The bryozoan larvae were found to contain both the bryostatins as well as a monoculture of symbiotic proteobacteria identified as *Endobugula sertula candidatus* (Haygood and Davidson, 1997; and Davidson and Haygood, 1999).

The present invention recognizes that organisms such as bacteria contain nucleic acid molecules that encode polypeptides that catalyze the synthesis of bioactive compounds, such as ecteinascidins.

As a non-limiting introduction to the breath of the present invention, the present invention includes several general and useful aspects, including:

1) a bacteria that at least in part synthesizes at least one ecteinascidin;
2) a preparation of bacteria from an *Ecteinascidia*, such as *Ecteinascidia turbinata*, that includes at least one activity involved in the synthesis of an ecteinascidin;
3) a composition including at least one nucleic acid molecule that encodes at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin;
4) a composition including a library of nucleic acid molecules of the present invention, where one or more of these nucleic acid molecules can be used in a combinatorial biosynthesis of ecteinascidins;
5) a composition including at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin, where preferably at least one polypeptide is derived from at least one marine organism;
6) a composition including a library of polypeptides of the present invention;

7) an antibody that binds with or specifically binds with a bacteria of the present invention;

8) an antibody that binds with or specifically binds with a polypeptide of the present invention;

9) a method of making an ecteinascidin using a bacteria of the present invention;

10) a composition made using a bacteria of the present invention;

11) a method of making a composition, including providing at least one nucleic acid of the present invention and synthesizing at least one ecteinascidin;

12) a composition made using a nucleic acid of the present invention;

13) a method of making a composition including providing at least one polypeptide of the present invention and synthesizing at least one ecteinascidin;

14) a composition made using a polypeptide of the present invention;

15) a method for identifying at least one nucleic acid molecule encoding at least one activity of involved in the synthesis of an ecteinascidin, comprising contacting a nucleic acid molecule of the present invention with a sample, and identifying nucleic acid molecules in said sample that hybridize with a nucleic acid of the present invention;

16) a nucleic acid molecule identified by a method of the present invention;

17) a composition comprising a library of nucleic acid molecules identified using a method of the present invention;

18) a method for identifying a bioactive compound including contacting a compound made or identified using a bacteria of the present invention and determining the bioactivity of the compound;

19) a compound identified by this method;

20) a method for identifying a bioactive compound including contacting a compound made or identified using a nucleic acid molecule of the present invention and determining the bioactivity of the compound;

21) a compound identified by this method;

22) a method for identifying a bioactive compound including contacting a compound made or identified using a polypeptide of the present invention and determining the bioactivity of said compound; and

23) a compound identified by this method.

These aspects of the invention, as well as others described herein, can be achieved by using the methods, articles of manufacture and compositions of matter described herein. To gain a full appreciation of the scope of the present invention, it will be further recognized that various aspects of the present invention can be combined to make desirable embodiments of the invention.

I. Bacteria That Synthesize An Ecteinascidin At Least In Part

The bacteria may be a member of any bacterial species, such as species belonging to the eubacteria or archaebacteria, that synthesize an ecteinascidin, at least in part. Preferably, the bacteria are marine bacteria, such as symbionts of the tunicate *Ecteinascidia turbinata*. The bacteria are isolated from their natural growth environment such that they are cultured in the absence of their host organism.

Isolation, purification, and culturing of bacteria can be accomplished using standard microbiological techniques as they are known in the art. Sterile technique can be employed to prevent the introduction of exogenous bacteria of fungi into the preparation. The bacteria of interest may be isolated from *E. turbinata* by disrupting or homogenizing a preparation of *E. turbinata*. Disruption or homogenization may be done by crushing, grinding, slicing, or rupturing the *E. turbinata*. This may be done manually, with or without

implements such as homogenizers, razor blades, scalpels, mortar and pestle, and the like, or may be done with any of a variety of machines such as, but not limited to, a blender or a motorized homogenizer.

The disruptate or homogenate may be filtered to remove large particles or debris from the preparation. In this case the filter has a pore size large enough to allow bacteria to pass through. Centrifugation may be used to sediment the bacteria. Preferably this is done after removing large particles from the disruptate or homogenate. After centrifugation, the supernatant is removed and the pellet containing the bacteria may be resuspended in a buffer or media.

The buffer or media to be used in resuspending the bacteria can be any media, but preferably is a media that promotes the growth of the bacteria. The media optionally contains a carbon source, a nitrogen source, salts, extracts such as yeast, beef, or liver extract, and/or vitamins. It may be desirable to include extracts of the *E. turbinata* host organism in the media. Such extracts may be taken from the disruptate or homogenate supernatant, or may be fractions of an *E. turbinata* disruptate or homogenate obtained by one or more centrifugations, column chromatography steps, precipitations, dialyses, or other fractionation or purification steps. Alternatively, samples of *E. turbinata* can be processed by sterilization, such as by autoclaving, and the resulting preparation used in microbiological medium by the optional inclusion of a gelling agent, such as pectin, agar or gelatin. Other nutrients can be added to such medium, such as vitamins, minerals and carbon and/or nitrogen sources. Optionally, an ocean environment can be created by adding autoclaved or filter sea-water, or artificial sea water (such as Instant Ocean) to such media. Substances that inhibit the growth of contaminants can be included in the medium. Such substances can be known antibiotics, or can be other substances found to inhibit the growth of contaminant, and can be ecteinascidins.

Cultures may be obtained by plating the bacteria on solid or semi-solid media or by propagating in liquid media and incubating the bacteria at a temperature, pressure and atmospheric conditions that allows the bacteria to propagate. Preferably, such culture conditions mirror or are similar to those that

the microorganism encounters in its native environment. Pure cultures may be obtained by a variety of art recognized methods, such as sequentially streaking colonies of bacteria from the culture, and/or by sequential dilution of bacterial cultures. Cultures enriched for a particular microorganism can be obtained by culturing a sample in a selective or differential microbiological medium such that the relative proportion of a particular microorganism is increased relative to other microorganisms. Such enrichment procedures can be used restoratively to provide for enriched cultures of varying purity.

The bacteria may be cultured on solid or semi-solid media, but once purified, the bacteria are preferably cultured in liquid media. When cultured in liquid media, the cultures may be shaken for aeration, or air may be injected into the culture through a tube to introduce oxygen into the culture, if such bacteria are facultatively aerobic or aerobic in metabolism. Preferably, the media used to culture such microorganisms mirrors or is related to the natural environment of the microorganism.

Preferably the bacteria are grown to optimal density for the production of ecteinascidins or ecteinascidin precursors, and the bacterial culture is harvested and centrifuged to sediment the bacteria. If the culture media is to be used for extraction of one or more ecteinascidins or ecteinascidin precursors, the supernatant is removed from the bacterial pellet. Purification of one or more ecteinascidins or ecteinascidin precursors may be performed through any methods known or developed in the future in the art, including combinations of any of the following: dialysis, precipitations, solvent partition, and chromatography, including high speed countercurrent chromatography (HSCCC), nonnal, and reverse phase chromatography. Other methods for purifying organic molecules are known in the art and may be employed as appropriate.

Activities of ecteinascidins may be assayed by any of the following art-recognized methods: An antitumor activity may be assayed *in vivo* using xenografts of human tumors in nude mice (see, for example, Veloti et al., 1977-1983 and U.S. Patent Numbers 5,256,663 and 5,089,273, herein incorporated by

reference) or *ex vivo* (Hakala, et al., 1987). An anti-cancer cell activity may be assayed using cell lines such as the P-388 murine leukemia cell line, the HT-29 human colon cancer cell line, and the A-549 human lung adenocarcinoma cell line (see, for example, U.S. Patent 5,478,932, and 5,089,273 hereby incorporated by reference and references disclosed therein.) A cytotoxic activity can be assayed by using the L1210 murine leukemia cells or the CV-1 monkey kidney cell line as disclosed in U.S. Patents 5,478,932 and 5,256,663, both hereby incorporated by reference. Assays to determine antibacterial and antiviral activities are also well known in the art and may be done as disclosed in U.S. Patent numbers 5,089,273 and 5,484,717, and references disclosed therein.

II. Nucleic Acid Molecules That Encode Polypeptides That Catalyze the Synthesis of An Ecteinascidin At Least In Part and Libraries of Such Nucleic Acid Molecules

The present invention includes a composition including at least one nucleic acid molecule, such as a substantially purified or purified nucleic acid molecule, that encodes at least a portion of at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin. Preferably, at least one nucleic acid molecule is derived from at least one marine organism. The nucleic acid molecules of the present invention can comprise the nucleic acid molecules disclosed herein, including PCR primers, portions thereof, and nucleic acid molecules that selectively hybridize with or have substantial identity with the nucleic acid molecules disclosed herein or portions thereof, or encode at least one conservative amino acid substitution relative to the disclosed sequences or portions thereof. A nucleic acid molecule of the present invention can be DNA, RNA, single stranded, double stranded or any combination thereof.

A nucleic acid molecule of the present invention preferably encodes at least a portion of a polypeptide involved in the synthesis of at least one ecteinascidin. Preferably, the polypeptide is at least a portion of a protein or complex that has non-ribosomal peptide transferase activity, that has O-methyl transferase activity, that has transaminase activity or that participates in the

synthesis or modification of one or more ecteinascidins. A nucleic acid molecule of the present invention preferably encodes at least a portion of at least one such active domain and can include at least one activity of such an active domain, preferably an activity that catalyzes at least one step in the synthesis of an ecteinascidin. Preferably, a nucleic acid molecule of the present invention is between about 10 bases or base pairs and about 500 Kb, between about 100 bases or base pairs and about 250 Kb or between about 500 bases or base pairs and about 100 Kb in length.

A nucleic acid molecule of the present invention can be derived from at least one marine organism. A marine organism can include any organism that can be found in a marine environment, either naturally or xenotypically. A marine organism can be a vertebrate, an invertebrate or a unicellular organism, such as a fungi, algae or bacteria. Preferably, a marine organism is an invertebrate, such as an *Ecteinascidia*, such as *Ecteinascidia turbinata*, or a unicellular organism, such as a bacteria, such as a bacteria associated with an *Ecteinascidia*, such as *Ecteinascidia turbinata*.

A nucleic acid of the present invention can also encode a fusion protein that includes a polypeptide of the present invention and a polypeptide of interest. A polypeptide of interest can be any polypeptide, but is preferably a detectable label, such as green fluorescent protein, or a sequence that aids in the purification of a polypeptide, such as FLAG. A nucleic acid that encodes a fusion protein can be made by operably linking a nucleic acid that encodes a polypeptide of interest with a nucleic acid that encodes a polypeptide of the present invention. The operable linking can be direct or indirect, such as in the case where a linker connects the polypeptide of the present invention with a polypeptide of interest. The nucleic acid molecule of the present invention and the nucleic acid that encodes a polypeptide of interest are preferably operably linked in frame such that an operable polypeptide of the present invention and an operable polypeptide of interest are translated from the nucleic acid, but that need not be the case.

Nucleic acid molecules of the present invention can be made using methods known in the art and described herein (see, Sambrook et al., supra (1989); Ausubel et al., supra (1998)). For example, nucleic acid molecules of the present invention can be identified and isolated using PCR methodologies, including RT-PCR, and sequenced using established methods such that their homologies can be determined. The ability of one nucleic acid molecule to hybridize with another can be determined through experimentation under a variety of stringencies, or can be estimated based on their length and G:C contents. Alterations of identified sequences can be made using routine methods, such as mutagenesis, RT-PCR or other PCR methods (See, Sambrook et al., 1989; Ausubel et al., 1998).

A nucleic acid molecule of the present invention can include at least one expression control sequence. Preferably, an expression control sequence is operably linked to a nucleic acid molecule such that the nucleic acid molecule can be expressed in an *in vivo* or *in vitro* transcription and/or translation system. The choice of expression control sequences is dependent upon the transcription system to be used. For example, if a prokaryotic organism such as *E. coli* is to be used to express a nucleic acid molecule, then at least one appropriate prokaryotic expression control sequence would be used. Likewise, if a eukaryotic organism is to be used to express a nucleic acid molecule, then at least one appropriate eukaryotic expression control sequence, such as the CMV promoter or LTRs would be used. Such nucleic acid molecules can be in any form, such as in a plasmid or in a linear form.

A nucleic acid molecule of the present invention can be provided with or without expression control sequences in a vector, such as a plasmid or a viral vector. Viral vectors can be chosen so that they are appropriate for a cell to be transfected, such as, for example, a phage, cosmid, retrovirus, vaccinia virus, adenovirus or adeno-associated virus. Viral vectors can introduce a nucleic acid molecule into a cell during its normal biological processes. Non-viral vectors can be used to introduce a nucleic acid molecule of the present invention into a host cell using methods known in the art, such as lipofection, cold calcium

chloride treatment or electroporation. The nucleic acid molecule in a cell can be extrachromosomal or be integrated into the genome of the cell. The host cell can be any appropriate host cell, such as a eukaryotic or prokaryotic cell. Preferably, a nucleic acid molecule of the present invention is expressed in the cell, but that is not a requirement of the invention. Preferably, the cell does not normally include a nucleic acid molecule of the present invention or express a polypeptide of the present invention, but that need not be the case. For example, a cell that expresses a relatively low amount of a polypeptide of the present invention can be made to express relatively higher amounts of a polypeptide once transfected with a nucleic acid of the present invention.

Cells that express a polypeptide of the present invention can be screened for and selected using a variety of methods, including those set forth in the present invention. For example, immunoassays, such as western blots, can be used to identify cell lysates that include a polypeptide of the present invention. In addition, immunocytochemistry can be used can be used to identify and localize a polypeptide of the present invention on or within a cell. Furthermore, *in situ* hybridization methods, such as FISH, can be used to identify and localize nucleic acid molecules within a cell and hybridization methods can be used to identify nucleic acid molecules, either DNA or RNA for cellular preparations. Cells or cell lysates can be screened for an activity using a variety of methods. For example, the ability of a cell or cell lysate to bind with a substrate or convert a substrate, including a detectably labeled substrate, can be used to detect a particular activity (see Haygood and Davidson, 1997).

Nucleic acid molecules of the present invention can encode peptides that have non-ribosomal peptide synthetase activity, have O-methyl transferase activity, have transaminase activity, or that participate in the synthesis or modification of one or more ecteinascidins. Nucleic acid molecules encoding peptides having such activity can be identified by making comparisons of nucleic acid sequence or translation amino acid sequences derived therefrom using methods known in the art, including BLAST comparisons. A nucleic acid molecule of the present invention can be expressed and the expression products

screened and confirmed for having such activity. In addition, nucleic acid molecules of the present invention can encode polypeptides that have other activities. Methods for screening such activities are known in the art.

The present invention also includes a library of nucleic acids of the present invention. A library of nucleic acids includes between about two, about four, about six, about eight, about ten, about thirty, about seventy, about one-hundred, about one-thousand, about ten-thousand, about one-hundred thousand or about one-million nucleic acid molecules and about three, about five, about seven, about twenty, about fifty, about five-hundred, about fifty-thousand, about five-hundred thousand and about ten million nucleic acid molecules. The members of such a library are preferably different nucleic acid molecules, but that need not be the case.

The nucleic acid molecules of the present invention can be used for a variety of applications, and therefore include but are not limited to, PCR primers, probes to identify similar sequences, and constructs for *in vitro* or *in vivo* transcription and/ or translation to make polypeptides of the present invention. The particular application of a nucleic acid molecule depends on a variety of factors, such as they are known in the art, include the length, strandedness (single stranded or double stranded and positive sense or negative sense), chemical characterization (such as DNA or RNA) or whether the nucleic acid molecule is detectably labeled or not.

III. Polypeptides That Catalyzes the Synthesis of An Ecteinascidin At Least in Part and Libraries Thereof

The present invention also includes a composition including at least one polypeptide or a portion thereof that catalyzes at least one step in the synthesis of at least one ecteinascidin, wherein the at least one polypeptide or a portion thereof is preferably derived from at least one marine organism.

A polypeptide of the present invention can be derived from at least one marine organism. A marine organism can include any organism that can be found in a marine environment, either naturally or xenotypically. A marine

organism can be a vertebrate, an invertebrate or a unicellular organism, such as a fungi, algae or bacteria. Preferably, a marine organism is an invertebrate, such as an *Ecteinascidia* such as *Ecteinascidia turbinata*, or a unicellular organism, such as a bacteria, such as a bacteria of the present invention.

The nucleic acid molecules of the present invention can be translated to provide polypeptides. These polypeptides can be substantially purified or purified and preferably have at least one activity of a polypeptide that participates in the synthesis of an ecteinascidin, such as but not limited to, a non-ribosomal peptide synthetase (NRPS), an O-methyl transferase, a transaminase, or any polypeptide that participates in the synthesis or modification of an ecteinascidin. Assays for non-ribosomal peptide synthetases are known in the art (see for example, Marahiel et al., 1997), and references disclosed therein). The activity of the polypeptide can be screened and confirmed using methods known in the art, later developed or described herein. For example, antibodies that bind with active portions or fragments of an activity can be used to identify appropriate polypeptides. Alternatively, substrates for an activity, such as substrates that are detectably labeled, can be used to detect the binding of a substrate to an activity or the conversion of a substrate to a product.

A polypeptide of the present invention can be of any length, but is preferably between about 10 amino acids and about 300,000 amino acids in length and more preferably between about 100 amino acids and about 100,000 amino acids in length or between about 1,000 amino acids and about 30,000 amino acids in length.

The polypeptide of the present invention can be made using recognized methods, such as by way of recombinant methods as they are known in the art (see, Sambrook et al., 1989; Ausubel et al., 1998) or by digesting proteins or polypeptides. For example, nucleic acid molecules encoding or suspected of encoding a polypeptide of the present invention can be cloned into expression vectors that are transformed into appropriate host cells where the nucleic acid molecules are expressed. The resulting polypeptides can be optionally purified and their activity confirmed using methods of the present invention or as they are

known in the art or later developed. Alternatively, the *in vivo* activity of polypeptides can be confirmed using methods of the present invention or as they are known in the art.

A polypeptide of the present invention can be provided *ex vivo* or within a cell. A polypeptide of the present invention can be expressed within a cell by transfecting a cell with a nucleic acid molecule that encodes a polypeptide of the present invention. The nucleic acid molecule of the present invention can be operably linked to expression control sequences appropriate for the cell such that the nucleic acid molecule of the present invention is expressed on or within the cell. The nucleic acid molecule can also encode a fusion protein such that the fusion protein is expressed on or within the cell. In this instance, a fusion protein that includes a detectable label as the polypeptide of interest can be used to confirm the presence or track the location of the fusion protein in the cell.

A polypeptide of the present invention can also be part of a fusion protein that includes a polypeptide of the present invention and a polypeptide of interest. A polypeptide of interest can be any polypeptide, but is preferably a detectable label, such as green fluorescent protein, or a sequence that aids in the purification of a polypeptide, such as FLAG. A fusion protein that includes a polypeptide of the present invention can be made from a nucleic acid that encodes a fusion protein can be made by operably linking a nucleic acid that encodes a polypeptide of interest with a nucleic acid that encodes a polypeptide of the present invention. The operable linking can be direct or indirect, such as in the case where a linker connects the polypeptide of the present invention with a polypeptide of interest. The nucleic acid molecule of the present invention and the nucleic acid that encodes a polypeptide of interest are preferably operably linked in frame such that an operable polypeptide of the present invention and an operable polypeptide of interest are translated from the nucleic acid, but that need not be the case. The present invention also includes such fusion proteins or libraries of such fusion proteins.

The present invention also includes a library of polypeptides of the present invention. A library of polypeptides of the present invention, including

fusion proteins, includes between about two, about four, about six, about eight, about ten, about thirty, about seventy, about one-hundred, about one-thousand, about ten-thousand, about one-hundred thousand or about one-million polypeptides and about three, about five, about seven, about twenty, about fifty, about five-hundred, about fifty-thousand, about five-hundred thousand and about ten million polypeptides. The members of such a library are preferably different polypeptides, but that need not be the case.

IV. Antibodies that Bind with a Bacteria of the Present Invention or a Polypeptide of the Present Invention

The present invention also includes antibodies that bind with or specifically bind with a bacterial or polypeptide of the present invention. The present invention also includes active fragments of such antibodies and methods of detecting bacteria of the present invention and polypeptides of the present invention. In such methods, the binding of an antibody to a target can be detected using direct or indirect detection methods, preferably using a detectable label, such as an enzyme, particle, or fluorescent moiety such as green fluorescent protein.

Bacteria

The present invention also includes antibodies that bind with or specifically bind with a bacteria of the present invention. Such antibodies can be polyclonal or monoclonal and can be made using methods known in the art (see, Harlowe and Lane, 1988). The antibodies can be whole antibodies of any class or subclass or mixture thereof, and can include active fragments of such antibodies, such as Fab fragments made using methods known in the art. The specificity of such antibodies can be screened and confirmed using assay formats known in the art, such as enzyme linked immunosorbent assays (ELISAs) or other appropriate immunoassay formats. Such antibodies are useful in detecting the presence and/or amount of such bacteria in a sample.

Polypeptides

The present invention also includes antibodies that bind with or specifically bind with a polypeptide of the present invention. Such antibodies can be polyclonal or monoclonal and can be made using methods known in the art (see, Harlowe and Lane, 1988). The antibodies can be whole antibodies of any class or subclass or mixture thereof, and can include active fragments of such antibodies, such as Fab fragments made using methods known in the art. The specificity of such antibodies can be screened and confirmed using assay formats known in the art, such as using enzyme linked immunosorbent assays (ELISAs) or other appropriate immunoassay formats. Such antibodies are useful in detecting the presence and/or amount of such polypeptides in a sample.

V. Method of Making Ecteinascidins Using Bacteria, Nucleic Acids or Polypeptides of the Present Invention and Compositions Made Thereby

The present invention also includes a method of making a composition including providing at least one bacteria, nucleic acid molecule or polypeptide of the present invention, and synthesizing at least one ecteinascidin or precursor thereof.

Bacteria

At least one species of bacteria of the present invention can be used to perform at least one step in the synthesis of an ecteinascidin. The ecteinascidin or ecteinascidin precursor or precursors synthesized by this method can be any previously known or unknown ecteinascidin, including novel ecteinascidins synthesized by modification of the biosynthesis pathway of a naturally occurring ecteinascidin, and including novel compounds synthesized using nucleic acids or polypeptides of the present invention that participate in ecteinascidin synthesis.

The bacteria may be a member of any bacterial species, such as species belonging to the eubacteria or archaebacteria, that synthesize an ecteinascidin, at least in part. Preferably, the bacteria are marine bacteria, such as symbionts of the tunicate *Ecteinascidia turbinata*. The bacteria are isolated from their natural

growth environment such that they are cultured in the absence of their host organism. Bacteria of the present invention may or may not be transformed with nucleic acids from other organisms, such as other bacterial species. Polypeptides encoded by such nucleic acids may participate in the synthesis of ecteinascidins, including ecteinascidins not naturally synthesized by the cultured bacteria, including modified forms of naturally occurring ecteinascidins.

Alternatively, the bacteria may be novel or previously identified bacteria that do not naturally synthesize ecteinascidins, but are transformed with nucleic acids of the present invention such that expression of the nucleic acids allow the bacteria synthesize one or more ecteinascidins or one or more ecteinascidin precursors. The bacteria may be of any bacterial species, such as species belonging to the eubacteria or archaebacteria. In addition, novel or previously identified bacteria, such as *Streptomyces lavendulae, Pseudomonas fluorescens* and *Myxococcus xanthus*, that synthesize compounds similar to ecteinascidins, may be transformed with nucleic acids of the present invention such that the bacteria produce novel compounds.

Isolation, purification, and culturing of bacteria can be accomplished using standard microbiological techniques as they are known in the art. Sterile technique can be employed to prevent the introduction of exogenous bacteria of fungi into the preparation. The bacteria of interest may be isolated from *E. turbinata* by disrupting or homogenizing a preparation of *E. turbinata*. Disruption or homogenization may be done by crushing, grinding, slicing, or rupturing the *E. turbinata*. This may be done manually, with or without implements such as homogenizers, razor blades, scalpels, mortar and pestle, and the like, or may be done with any of a variety of machines such as, but not limited to, a blender or a motorized homogenizer.

The disruptate or homogenate may be filtered to remove large particles or debris from the preparation. In this case the filter has a pore size large enough to allow bacteria to pass through. Centrifugation may be used to sediment the bacteria. Preferably this is done after removing large particles from the disruptate

or homogenate. After centrifugation, the supernatant is removed and the pellet containing the bacteria may be resuspended in a buffer or media.

The buffer or media to be used in resuspending the bacteria can be any media, but preferably is a media that promotes the growth of the bacteria. The media optionally contains a carbon source, a nitrogen source, salts, extracts such as yeast, beef, or liver extract, and/or vitamins. It may be desirable to include extracts of the *E. turbinata* host organism in the media. Such extracts may be taken from the disruptate or homogenate supernatant, or may be fractions of an *E. turbinata* disruptate or homogenate obtained by one or more centrifugations, column chromatography steps, precipitations, dialyses, or other fractionation or purification steps. Alternatively, samples of E. turbinata can be processed by sterilization, such as by autoclaving, and the resulting preparation used in microbiological medium by the optional inclusion of a gelling agent, such as pectin, agar or gelatin. Other nutrients can be added to such medium, such as vitamins, minerals and carbon and/or nitrogen sources. Optionally, an ocean environment can be created by adding autoclaved or filter sea-water, or artificial sea water (such as Instant Ocean) to such media.

Cultures may be obtained by plating the bacteria on solid or semi-solid media or by propagating the bacteria in liquid media and incubating the bacteria at a temperature, pressure and atmospheric conditions that allows the bacteria to propagate. Such culture conditions may mirror or be similar to those that the microorganism encounters in its native environment.

Pure cultures may be obtained by a variety of art recognized methods, such as sequentially streaking colonies of bacteria from the culture. Cultures enriched for a particular microorganism can be obtained by culturing a sample in a selective or differential microbiological medium such that the relative proportion of a particular microorganism is increased relative to other microorganisms. Such enrichment procedures can be used restoratively to provide for enriched cultures of varying purity.

The bacteria may be cultured on solid or semi-solid media or in liquid media, but once purified, the bacteria are preferably cultured in liquid media. If

the medium is a solid or semi-solid medium, a gelling agent, such as pectin, agar or gelatin is included. The medium optionally includes a nitrogen source, salts, extracts such as yeast, beef, or liver extract, and/or vitamins. It may be desirable to include extracts of the *E. turbinata* host organism in the media. Such extracts may be taken from the disruptate or homogenate supernatant, or may be fractions of an *E. turbinata* disruptate or homogenate obtained by one or more centrifugations, column chromatography steps, precipitations, dialyses, or other fractionation or purification steps. Alternatively, samples of *E. turbinata* can be processed by sterilization, such as by autoclaving, and the resulting preparation used in microbiological medium. Optionally, an ocean environment can be created by adding autoclaved or filter sea-water, or artificial sea water (such as Instant Ocean) to such media. When cultured in liquid media, the cultures may be shaken for aeration, or air may be injected into the culture through one or more tubes to introduce oxygen into the culture, if such bacteria are facultatively aerobic or aerobic in metabolism. Preferably, the media used to culture such microorganisms mirrors or is related to the natural environment of the microorganism.

For production of ecteinascidins, bacteria may be cultured under appropriate conditions of temperature, salt, nutrients, pH, and aeration. The growth media preferably contains compounds that promote optimal growth of the bacteria, such as but not limited to, salts, sugars, amino acids, vitamins, extracts, and/or buffering agents. The bacteria may be cultured under fermentation conditions such that they may be grown to high density. In fermentative growth, it may be desirable to add compounds, such as sodium hydroxide, to stabilize the pH of the media as the cells grow. Alternatively, media may be exchanged during the period of culturing such that the growth media retains optimal levels of pH and nutrients.

Precursors of ecteinascidins, such as but not limited to amino acids, may be added to the bacterial media. Ecteinascidin precursors, and other compounds such as nutrients, may be added to the culture one or more times before and/or after inoculation of the media with bacteria.

Culture media may be harvested during the period of growth for isolation of ecteinascidins. Alternatively, the bacteria, the media, or both may be harvested at the end of the growth period.

Ecteinascidins or ecteinascidin precursors may be isolated and purified or partially purified from the media, bacteria, or both using appropriate methods as they are known or developed in the art. Such methods may include one or more steps of one or more of the following: centrifugation, phase separation, precipitation, chromatography, dialysis. The purity of the ecteinascidins or ecteinascidin precursors may be evaluated by chemical or biophysical means, such as 1D or 2D NMR and their activity assayed by antitumor, anticancer, cytotoxic, antibacterial, or antiviral assays as they are known in the art, including those disclosed in U.S. patents 5,256,663, 5,478,932,5,484,717, and 5,089,273.

The present invention also includes compounds made or identified using the present invention. For example, the present invention includes ecteinascidins made using at least one method of the present invention. A compound made or identified using a method of the present invention can be a novel or non-novel compound. For example, a compound of the present invention can include a compound that was not or was not novel on the date of the filing of the present application, or one year or six months prior to the filing date of the present application.

A compound of the present invention can be provided with at least one pharmaceutically acceptable carrier as they are known in the art and discussed herein. Such pharmaceutically acceptable carriers are known in the art and are disclosed herein. A compound of the present invention can also be a pharmaceutical composition.

Nucleic Acid Molecules and Polypeptides

At least one nucleic acid molecule of the present invention or at least one polypeptide of the present invention can be expressed and used in a system to synthesize an ecteinascidin. The ecteinascidin or precursors thereof can be previously known or unknown ecteinascidin or precursors thereof.

The present invention utilizes at least one nucleic acid molecule of the present invention and/or at least one polypeptide of the present invention in such methods to make known or novel ecteinascidins or related compounds, including Et-743. The present invention can utilize at least one nucleic acid molecule of the present invention and/or at least one polypeptide of the present invention alone or in combination with other genes. These polypeptides and genes can be known or later developed and can be derived from marine, aquatic or terrestrial organisms.

For example, some methods of synthesizing ecteinascidins may provide cassettes that include a gene complex that encodes at least one ecteinascidin synthesis activity, such as a non-ribosomal peptide synthetase activity, either in whole or in part. Nucleic acid molecules of the present invention can be inserted into such cassettes randomly or non-randomly, including replacing identified genes. When nucleic acid molecules of the present invention are inserted non-randomly into such cassettes, they can be inserted in-frame to replace existing genes that encode polypeptides that have functions similar to the polypeptide encoded by a nucleic acid of the present invention. The nucleic acid molecules of the present invention are thus expressed as polypeptides of the present invention, which can act as part of a gene complex to produce known or novel compounds, such as ecteinascidins.

Cells or extracts thereof (such as substantially purified extracts) that include one or more of the nucleic acid molecules of the present invention or one or more polypeptides of the present invention can be used to synthesize a wide variety of ecteinascidins including Et-743. Such cells or extracts thereof can be contacted with a variety of compounds, including substrates for an ecteinascidin synthesis activity, particularly an ecteinascidin synthesis activity present in the cells or extracts thereof. Polypeptides expressed from nucleic acids of the present invention can act on these compounds in order to make a wide variety of ecteinascidins including Et-743. In one aspect of the present invention, more than one cell and/or extract thereof can be used in combination or sequentially

such that the products made by combination of cells or extracts can be determined and its activity confirmed.

Alternatively, polypeptides expressed from nucleic acids of the present invention can be produced and purified or substantially purified for use in the manufacture of one or more ecteinascidins, one or more ecteinascidin precursors, or one or more ecteinascidin related compounds, including novel compounds. The polypeptides may be produced by expression of the nucleic acids of the present invention in any suitable system, including *in vitro* transcription and translation systems and *in vivo* expression systems. For example, a polypeptide of the present invention can be expressed within a cell by transfecting a cell with a nucleic acid molecule that encodes a polypeptide of the present invention. The nucleic acid molecule of the present invention can be operably linked to expression control sequences appropriate for the cell such that the polypeptide of the present invention is expressed on or within the cell, or the polypeptide of the present invention that is expressed within a cell is secreted from the cell. Cells for expression of a polypeptide of the present invention include but are not limited to bacterial cells, such as *E. coli* or *Bacillus* cells, yeast cells, such as *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Pichia metholica*, eukaryotic cells, such as plant cells, insect cells, or mammalian cells including without limitation COS-1 cells (ATCC No. CRL 1650), COS-7 cells (ATCC No. CRL 1651), BHK cells (ATCC No. CRL1632), BHK 570 cells (ATCC No. CRL 10314), 293 cells (ATCC No. CRL 1573) and CHO-K1 cells (ATCC No. CCL 61). Transformed or transfected cells can be cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the particular host cells, and preferably, for the selection of cells containing the exogenously added nucleic acid.

A polypeptide of the present invention may be purified from the cell culture using well-established methods of protein purification, preferably including affinity purification (see, for example, Scopes, 1994; 1988).

The present invention also includes compounds made or identified using the present invention. For example, the present invention includes ecteinascidins

made using at least one method of the present invention. A compound made or identified using a method of the present invention can be a novel or non-novel compound. For example, a compound of the present invention can include a compound that was or was not novel on the date of the filing of the present application, or one year or six months prior to the filing date of the present application.

A compound of the present invention can be provided with at least one pharmaceutically acceptable carrier as they are known in the art and discussed herein. Such pharmaceutically acceptable carriers are known in the art and are disclosed herein. A compound of the present invention can also be a pharmaceutical composition.

VI. Method of Identifying Nucleic Acid Molecules, Nucleic Acid Molecules Identified Thereby and Libraries of Nucleic Acid Molecules

The present invention also includes a method for identifying at least one nucleic acid molecule encoding at least one ecteinascidin synthesis activity including contacting a nucleic acid molecule of the present invention with a sample, and identifying nucleic acid molecules in said sample that hybridize with said nucleic acid molecule of the present invention. This aspect of the present invention utilizes nucleic acid molecules of the present invention as probes or PCR primers in order to identify nucleic acid molecules that have or are expected to encode polypeptides that have ecteinascidin synthesis activity.

Samples for use in the present invention can be from any source that can include a nucleic acid molecule, but preferably include samples from an environmental sample, such as the marine environment. The samples can include marine organisms, including invertebrates or vertebrates or any other marine organism. Preferably, the sample includes single celled organisms, such as bacteria. More preferably, the sample includes samples that are expected to contain an ecteinascidin. Such samples include, for example, *Ecteinascidia* species, including *Ecteinascidia turbinata.*

When used as probes, nucleic acid molecules of the present invention can be detectably labeled and contacted with a sample. Nucleic acid molecules that bind with the nucleic acid of the present invention can be identified, cloned and sequenced using methods known in the art. The identified nucleic acid molecules can be operably linked to expression control sequences such that a polypeptide encoded by the identified nucleic acid molecule can be made and characterized.

When used as PCR primers, the nucleic acid molecules of the present invention can be used to amplify nucleic molecules in a sample. One or more than one of the primers used in one or more amplification reactions can be a nucleic acid molecule of the present invention. The amplified nucleic acid molecules are presumptively derived from a gene encoding an ecteinascidin synthesis activity. The amplified nucleic acid molecules can be identified, cloned and sequenced using methods known in the art. The amplified nucleic acid molecules can be operably linked to expression control sequences such that a polypeptide encoded by the amplified nucleic acid molecule can be made and characterized.

The nucleic acid molecules of the present invention can also be used to identify nucleic acid molecules that are upstream or downstream from a targeted segment of a gene encoding an ecteinascidin synthesis activity. A nucleic acid molecule that encodes a conserved region of a gene encoding an ecteinascidin synthesis activity. A nucleic acid molecule of the present invention can be used in primer extension or inverse PCR methods such that upstream or downstream segments from the point of hybridization are identified. These extended segments can be identified, cloned and sequenced using methods known in the art. The extended segments can be operably linked to expression control sequences such that a polypeptide encoded by the extended segment can be made and characterized.

Whether the nucleic acid molecule of the present invention is used as a probe, primer or PCR primer, the methods of the present invention identify nucleic acid molecules that presumptively encode at least a portion of a gene

encoding an ecteinascidin synthesis activity. Any of these processes can be used alone, in combination or restoratively to identify at least portions of genes encoding an ecteinascidin synthesis activity in a sample.

The present invention also includes nucleic acid molecules identified by the present invention. The identified nucleic acid molecules can include expression control sequences operably linked to the identified nucleic acid molecules. Such constructs can be used to make polypeptides encoded by the identified nucleic acid molecules and the polypeptides can be characterized as to a variety of structures and functions, particularly structures and functions associated with genes encoding an ecteinascidin synthesis activity. The present invention includes a library of nucleic acids, cells or polypeptides identified by the present invention.

VII. Method of Identifying a Bioactive Compound, Bioactive Compounds, and Therapeutic Compositions

The present invention also includes a method for identifying a bioactive compound, such as those made or identified using a bacteria, nucleic acid molecule or polypeptide of the present invention, including contacting a compound made or identified by the present invention with at least one *in vitro*, *ex vivo* or *in vivo* assay system and determining the bioactivity of said compound. The present invention includes bioactive compounds identified using this method. The identified bioactive compounds can be provided in a pharmaceutically acceptable carrier and can be a pharmaceutical compound.

In vitro, *ex vivo* and *in vivo* systems used in the present invention are preferably those known in the art for a bioactivity to be identified. The assay chosen to be used in this method is related to a bioactivity that is being screened for. Preferred systems include those that determine at least one ecteinascidin activity. For example, *in vitro* systems (systems that do not use whole organisms or whole cells) and *ex vivo* systems (systems that use whole cells or portions of cells) for the identification of ecteinascidin activity are known in the art, including those disclosed in U.S. patents 5,256,663, 5,478,932,5,484,717, and

5,089,273. *In vivo* systems (systems that use whole organisms or tissues or organs derived therefrom) are also known in the art (see, for example, Valoti et al., 1998). Compounds are identified using these methods as having a desired bioactivity. Compounds identified by these methods are bioactive compounds that have at least one bioactivity.

Screening of compounds for activities

The following assays can be performed to confirm the bioactivity of a compound:

a) antimicrobial effect on *S. aureus* by placing a compound on a paper disk and determining the ability of the compound to inhibit the growth of the *S. aureus* (Benson, 1994). The results of this assay establish the toxicity of the compound towards Gram-positive bacteria.

b) antimicrobial effect on *E. coli* by placing a compound on a paper disk and determining the ability of the compound to inhibit the growth of the *E. coli* (Benson, 1994). The results of this assay establish the toxicity of a compound towards Gram-negative bacteria.

c) antimicrobial effect on *Candida albicans* by placing a compound on a paper disk and determining the ability of the extract to inhibit the growth of *Candida albicans* (Benson, 1994). The results of this study establish the toxicity of compounds towards yeasts and fungi.

d) inhibition of the growth of cancer cells by contacting a compound with the National Cancer Institute's (NCI) cell line screen (approximately sixty cell lines) against up to fifty-one cancer cell types *in vitro* (Boyd et al., 1995). These results provide an activity profile for the compound. The activity profiles of an extract can be compared to the activity profiles of other samples in the NCI database of activity profiles. Similar activity profiles of different extracts, including known extracts with known modes of action, strongly suggests that the samples have similar modes of action. A novel activity profile strongly suggests that the compound has a novel mechanism of action.

e) cytotoxic activity of a compound can be determined by a variety of methods, including inhibition of brine shrimp by contacting twenty-four hour old brine shrimp nauplii for twenty-four hours with an compound and observing the inhibition of the activity or viability of the brine shrimp. The results of this assay establish the cytotoxicity of the compound towards whole organisms.

Compounds identified as having a bioactivity have presumptive therapeutic activity. Such therapeutic activity and related pharmacological parameters can be confirmed using the methods discussed herein.

Pharmacology and toxicity of bioactive compounds and bioactivities

The structure of a bioactive compound or bioactivity can be determined or confirmed by methods known in the art, such as mass spectroscopy. For bioactive compounds and bioactivities stored for extended periods of time under a variety of conditions, the structure, activity and potency thereof can be confirmed.

Identified bioactive compounds and bioactivities can be evaluated for a particular activity using recognized methods and those disclosed herein. For example, if an identified bioactive compound or bioactivity is found to have anticancer cell activity *in vitro*, then the bioactive compound or bioactivity would have presumptive pharmacological properties as a chemotherapeutic to treat cancer. Such nexuses are known in the art for several disease states, and more are expected to be discovered over time. Based on such nexuses, appropriate confirmatory *in vitro* and *in vivo* models of pharmacological activity and toxicology can be selected and appropriate tests performed. The methods described herein can also be used to assess pharmacological selectivity and specificity, and toxicity.

Identified bioactive compounds and bioactivities can be evaluated for toxicological effects using known methods (see, Lu, 1985; U.S. Patent Nos; 5,196,313 to Culbreth (issued March 23, 1993) and 5,567,952 to Benet (issued October 22, 1996)). For example, toxicology of a bioactive compound or bioactivity can be established by determining *in vitro* toxicity towards a cell line,

such as a mammalian, for example human, cell line. Bioactive compounds and bioactivities can be treated with, for example, tissue extracts, such as preparations of liver, such as microsomal preparations, to determine increased or decreased toxicological properties of the bioactive compound of bioactivity after being metabolized by a whole organism. The results of these types of studies are predictive of toxicological properties of chemicals in animals, such as mammals, including humans.

Alternatively, or in addition to these *in vitro* studies, the toxicological properties of a bioactive compound or bioactivity in an animal model, such as mice, rats, rabbits, dogs or monkeys, can be determined using established methods (see, Lu, 1985; and Creasey, 1979). Depending on the toxicity, target organ, tissue, locus and presumptive mechanism of the bioactive compound or bioactivity, the skilled artisan would not be burdened to determine appropriate doses, LD_{50} values, routes of administration and regimes that would be appropriate to determine the toxicological properties of the bioactive compound or bioactivity. In addition to animal models, human clinical trials can be performed following established procedures, such as those set forth by the United States Food and Drug Administration (USFDA) or equivalents of other governments. These toxicity studies provide the basis for determining the efficacy of a bioactive compound or bioactivity *in vivo*.

Efficacy of bioactive compounds and bioactivities

Efficacy of a bioactive compound or bioactivity can be established using several art recognized methods, such as *in vitro* methods, animal models or human clinical trials (see, Creasey, 1979). Recognized in vitro models exist for several diseases or conditions. For example, the ability of a compound or composition to extend the life-span of HIV-infected cells in vitro is recognized as an acceptable model to identify chemicals expected to be efficacious to treat HIV infection or AIDS (see, Daluge et al., 1995). Furthermore, the ability of cyclosporin A (CsA) to prevent proliferation of T-cells in vitro has been established as an acceptable model to identify chemicals expected to be efficacious as immunosuppressants (see, Suthanthiran et al., 1996). For nearly

every class of therapeutic, disease or condition, an acceptable *in vitro* or animal model is available. In addition, these *in vitro* methods can use tissue extracts, such as preparations of liver, such as microsomal preparations, to provide a reliable indication of the effects of metabolism on a bioactive compound or bioactivity. Similarly, acceptable animal models can be used to establish efficacy of bioactive compounds and bioactivities to treat various diseases or conditions. For example, the rabbit knee is an accepted model for testing agents for efficacy in treating arthritis (see, Shaw and Lacy, 1973). Hydrocortisone, which is approved for use in humans to treat arthritis, is efficacious in this model which confirms the validity of this model (see, McDonough, 1982). When choosing an appropriate model to determine efficacy of bioactive compounds and bioactivities, the skilled artisan can be guided by the state of the art to choose an appropriate model, doses and route of administration, regime and endpoint and as such would not be unduly burdened.

In addition to animal models, human clinical trials can be used to determine the efficacy of bioactive compounds and bioactivities. The USFDA, or equivalent governmental agencies, have established procedures for such studies.

Selectivity of bioactive compounds and bioactivities

The *in vitro* and *in vivo* methods described above also establish the selectivity of a candidate modulator. It is recognized that chemicals can modulate a wide variety of biological processes or can be selective. Panels of cells as they are known in the art can be used to determine the specificity of the a bioactive compound or bioactivity (WO 98/13353 to Whitney et al., published April 2, 1998). Selectivity is evident, for example, in the field of chemotherapy, where the selectivity of a chemical to be toxic towards cancerous cells, but not towards non-cancerous cells, is obviously desirable. Selective modulators are preferable because they have fewer side effects in the clinical setting. The selectivity of a bioactive compound or bioactivity can be established *in vitro* by testing the toxicity and effect of a bioactive compound or bioactivity can be established *in vitro* by testing the toxicity and effect of a bioactive compound or

bioactivity on a plurality of cell lines that exhibit a variety of cellular pathways and sensitivities. The data obtained form these *in vitro* toxicity studies can be extended to animal model studies, including human clinical trials, to determine toxicity, efficacy and selectivity of a bioactive compound or bioactivity.

The selectivity, specificity and toxicology, as well as the general pharmacology, of a bioactive compound or bioactivity can be often improved by generating additional test chemicals based on the structure/property relationship of a bioactive compound or bioactivity originally identified as having activity. Bioactive compounds and bioactivities can be modified to improve various properties, such as affinity, life-time in blood, toxicology, specificity and membrane permeability. Such refined bioactive compounds and bioactivities can be subjected to additional assays as they are known in the art or described herein. Methods for generating and analyzing such compounds or compositions are known in the art, such as U.S. Patent No. 5,574,656 to Agrafiotis et al.

Pharmaceutical compositions

The present invention also encompasses a bioactive compound or bioactivity in a pharmaceutical composition comprising a pharmaceutically acceptable carrier prepared for storage and preferably subsequent administration, which have a pharmaceutically effective amount of the bioactive compound or bioactivity in a pharmaceutically acceptable carrier or diluent. Acceptable carriers or diluents for therapeutic use are well known in the pharmaceutical art, and are described, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co., (A.R. Gennaro edit. (1985)). Preservatives, stabilizers, dyes and even flavoring agents can be provided in the pharmaceutical composition. For example, sodium benzoate, sorbic acid and esters of p-hydroxybenzoic acid can be added as preservatives. In addition, antioxidants and suspending agents can be used.

The bioactive compounds and bioactivities of the present invention can be formulated and used as tablets, capsules or elixirs for oral administration; suppositories for rectal administration; sterile solutions, suspensions or injectable administration; and the like. Injectables can be prepared in

conventional forms either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. Suitable excipients are, for example, water, saline, dextrose, mannitol, lactose, lecithin, albumin, sodium glutamate, cysteine hydrochloride and the like. In addition, if desired, the injectable pharmaceutical compositions can contain minor amounts of nontoxic auxiliary substances, such as wetting agents, pH buffering agents and the like. If desired, absorption enhancing preparation, such as liposomes, can be used.

The pharmaceutically effective amount of a bioactive compound or bioactivity required as a dose will depend on the route of administration, the type of animal or patient being treated, and the physical characteristics of the specific animal under consideration. The dose can be tailored to achieve a desired effect, but will depend on such factors as weight, diet, concurrent medication and other factors which those skilled in the medical arts will recognize. In practicing the methods of the present invention, the pharmaceutical compositions can be used alone or in combination with one another, or in combination with other therapeutic or diagnostic agents. These products can be utilized *in vivo*, preferably in a mammalian patient, preferably in a human, or in vitro. In employing them *in vivo*, the pharmaceutical compositions can be administered to the patient in a variety of ways, including parenterally, intravenously, subcutaneously, intramuscularly, colonically, rectally, nasally or intraperitoneally, employing a variety of dosage forms. Such methods can also be used in testing the activity of bioactive compounds or bioactivities *in vivo*.

As will be readily apparent to one skilled in the art, the useful *in vivo* dosage to be administered and the particular mode of administration will vary depending upon the age, weight and type of patient being treated, the particular pharmaceutical composition employed, and the specific use for which the pharmaceutical composition is employed. The determination of effective dosage levels, that is the dose levels necessary to achieve the desired result, can be accomplished by one skilled in the art using routine methods as discussed above. Typically, human clinical applications of products are commenced at lower

dosage levels, with dosage level being increased until the desired effect is achieved. Alternatively, acceptable *in vitro* studies can be used to establish useful doses and routes of administration of the bioactive compounds and bioactivities.

In non-human animal studies, applications of the pharmaceutical compositions are commenced at higher dose levels, with the dosage being decreased until the desired effect is no longer achieved or adverse side effects are reduced or disappear. The dosage for the bioactive compounds and bioactivities of the present invention can range broadly depending upon the desired affects, the therapeutic indication, route of administration and purity and activity of the bioactive compound or bioactivity. Typically, dosages can be between about 1 ng/kg and about 10 mg/kg, preferably between about 10 ng/kg and about 1 mg/kg, more preferably between about 100 ng/kg and about 100 micrograms/kg, and most preferably between about 1 microgram/kg and about 10 micrograms/kg.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition (see, Fingle et al., in The Pharmacological Basis of Therapeutics (1975)). It should be noted that the attending physician would know how to and when to terminate, interrupt or adjust administration due to toxicity, organ dysfunction or other adverse effects. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate. The magnitude of an administrated does in the management of the disorder of interest will vary with the severity of the condition to be treated and to the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight and response of the individual patient, including those for veterinary applications.

Depending on the specific conditions being treated, such pharmaceutical compositions can be formulated and administered systemically or locally. Techniques for formation and administration can be found in Remington's

Pharmaceutical Sciences, 18th Ed., Mack Publishing Co., Easton, PA (1990). Suitable routes of administration can include oral, nasal, rectal, transdermal, otic, ocular, vaginal, transmucosal or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections.

For injection, the pharmaceutical compositions of the present invention can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution or physiological saline buffer. For such transmucosal administration, penetrans appropriate to the barrier to be permeated are used in the formulation. Such penetrans are generally known in the art. Use of pharmaceutically acceptable carriers to formulate the pharmaceutical compositions herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular, those formulation as solutions, can be administered parenterally, such as by intravenous injection. The pharmaceutical compositions can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administrations. Such carriers enable the bioactive compounds and bioactivities of the invention to be formulated as tables, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Substantially all molecules present in an aqueous solution at the time of liposome formation are incorporated into or within the liposomes thus formed. The liposomal contents are both protected from the external micro-environment and, because liposomes fuse will cell membranes, are efficiently delivered into

the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules can be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. Determination of the effective amount of a pharmaceutical composition is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. In addition to the active ingredients, these pharmaceutical compositions can contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active chemicals into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tables, dragees, capsules or solutions. The pharmaceutical compositions of the present invention can be manufactured in a manner that is itself known, for example by means of conventional mixing, dissolving, granulating, dragee-making, emulsifying, encapsulating, entrapping or lyophilizing processes. Pharmaceutical formulations for parenteral administration include aqueous solutions of active chemicals in water-soluble form.

Additionally, suspensions of the active chemicals may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides or liposomes. Aqueous injection suspensions may contain substances what increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol or dextran. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the chemicals to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions for oral use can be obtained by combining the active chemicals with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tables or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol or sorbitol; cellulose

preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose and/or polyvinylpyrrolidone. If desired, disintegrating agents can be added, such as the cross-linked polyvinyl pyrolidone, agar, alginic acid or a salt thereof such as sodium alginate. Dragee cores can be provided with suitable coatings. Dyes or pigments can be added to the tablets or dragee coatings for identification or to characterize different combinations of active doses.

The bioactive compounds and bioactivities of the present invention, and pharmaceutical compositions that include such bioactive compounds and bioactivities are useful for treating a variety of ailments in a patient, including a human. As set forth in the Examples, the bioactive compounds and bioactivities of the present invention have antibacterial, antimicrobial, antiviral, anticancer cell, antitumor and cytotoxic activity. A patient in need of such treatment can be provided a bioactive compound or bioactivity of the present invention, preferably in a pharmacological composition in an effective amount to reduce the number or growth rate of bacteria, microbes, cancer cells or tumor cells in said patent, or to reduce the infectivity of viruses in said patient. The amount, dosage, route of administration, regime and endpoint can all be determined using the procedures described herein.

The invention will now be described by the following non-limiting examples.

Example 1

Identification of Bacteria Associated With *Ecteinascidia turbinata* by *in situ* Hybridization

Larvae of *E. turbinata* for *in situ* hybridization are fixed in 4% paraformaldehyde-0.1 M MOPS (pH 7.4)-0.5 M NaCl. Fixed larvae are either used immediately or stored in 70% ethanol at -20°C. The fixed larvae are treated with proteinase K (Boehringer Mannheim), refixed in 4% paraformaldehyde, treated with acetic anhydride in 0.1 M triethanolamine, washed with buffer, and

incubated in hybridization buffer (5X SSC, 0.1% Tween 20, 5 mM EDTA, and 1X Denhardt's solution) at 55°C for one hour. Fluorescein-labeled probes are added in fresh hybridization buffer at a concentration of 2 ng per microliter. The samples are heated to 55°C for five to ten minutes and then incubated at 43°C for 4 to 6 hours. Each wash is conducted twice for 15 minutes as follows: high stringency wash 1 is 10% formamide, 5X SSC, and 1% sodium dodecyl sulfate, used at 47°C, wash solution 2 is 10% formamide, 2X SSC, and 1% sodium dodecyl sulfate, used at 37°C, and wash solution 3 is 10% formamide, 2X SSC, and 0.1% Tween 20, used at room temperature. The larvae are suspended in Vectasheild (Vector) mounting media and viewed under epifluorescence. The probes used for hybridization are designed to hybridize to regions of the SSU rRNA gene (Haygood and Davidson, 1997) and include the universal eubacterial probe EUB 338 (SEQ ID NO:4) (Amann et al., 1997), the *E. palauensis* SSU rRNA probe 1247r (5'-AAGTAGCGGCCCTTTGTC-3') (SEQ ID NO:1), and the *E. sertula* probe Bn1253r (SEQ ID NO:9) (Haygood and Davidson, 1997).

Example 2

Extraction of Nucleic Acid Molecules from Samples of *E. turbinata* that Include Bacteria

Unless otherwise noted, molecular biology procedures used are standard techniques used in the field (Sambrook et al., 1989; Ausubel et al., 1998).

DNA extraction

Samples of either adult or larval *E. turbinata* were obtained from the waters of the Florida Keys. Free-swimming larvae were collected in the lab from *E. turbinata* adult colonies. Larval DNA was extracted as previously described for *B. neritina* (Haygood and Davidson, 1997). Larvae were concentrated in 1.5 milliliter microcentrifuge tubes (approximately 25 mg of larvae per tube) by gentle centrifugation and then rinsed four times in filtered seawater (0.2 micrometer pore size filter) to minimize the contaminating seawater bacteria. Excess water was removed and the pellets were frozen at -80°C for later use.

DNA was extracted from the *E. turbinata* larvae using a QIAamp Tissue Kit as directed by the manufacturer (Qiagen Inc., Valencia, CA).

Example 3

Identification of Genes in Samples of *E. turbinata* that Include Bacteria

Primer design

Degenerate PCR primers are designed based on adenylation domains of non-ribosomal peptide synthetase (NRPS) genes. PCR primers are designed as follows: 1) using published degenerate sequences for the A-domains, and 2) using published saframycin MX1 sequences as a basis for more specific primers. The adenylation domain is selected as the first target because it is the most highly conserved region of the NRPS genes. In addition, it is the loading domain for amino acids and thus serves as the basis for amino acid selectivity in the NRPS genes. Recently, it has been reported that specific amino acids within this domain can be decoded to predict which amino acid is encoded by the domain. Therefore, comparison of isolated A-domain sequences with those in published databases will provide added evidence that the genes are encoding enzymes in the ecteinascidin pathway.

PCR conditions for amplification of DNA from E. turbinata

The PCR conditions for the initial amplification of the genes from *E. turbinata* DNA can be as follows. A total reaction volume of 50 microliters contains approximately 100 ng of *E. turbinata* DNA (either adult or larval), 1 micromolar of each of two degenerate primers, and Taq polymerase and buffer (Boehringer Mannheim Corp., Indianapolis, IN). A PCR protocol is optimized for the degenerate NRPS primers. The cycle conditions start with a "touch down" sequence, which lowers the annealing temperature from 60 to 40°C at a rate of 2°C per cycle (11 cycles), and were then maintains at 40°C for a total of 51 cycles. Cycle steps are as follows: denaturation (94°C; 1 minute), annealing (60°C to 40°C; 2 minutes), and extension (72°C; 1 minute).

DNA cloning

PCR reactions are electrophoresed on 0.8% agarose gel and visualized with ethidium bromide. PCR products of the expected length are cloned using a TOPO TA Cloning kit into the Invitrogen pCR®2.1-TOPO vector as described by the manufacturer (Invitrogen Corp., Carlsbad, CA). Recombinant clones containing insert DNA are then sequenced using standard protocols. Since NRPS enzymes can be modular, clones from the degenerate PCR primers can represent a pool of fragments from different NRPS domains.

DNA Sequencing

Plasmid DNA for sequencing is prepared using the Qiagen QIAprep Spin Miniprep Kit (Qiagen, Inc., Valencia, CA). All sequencing is performed with an ABI automated sequencer (model 373A) by using a PRISM Ready Reaction DyeDeoxy terminator cycle sequencing kit as recommended by the manufacturer (Perkin-Elmer). Cloned genes are sequenced using primers directed against the cloning vector, pCR®2.1-TOPO (Invitrogen, Carlsbad, CA).

Example 4

Sequences of Ecteinascidin Synthesis Activity from *E. turbinata* that Include Bacteria

Cloned DNA sequences may be identified by using the BLAST (basic local alignment search tool) server of the National Center for Biotechnology Information accessed over the Internet (Altschul et al., 1997). Clones that are identified as having homology to known NRPS genes can be used to probe libraries, such as cosmid libraries, made from genomic DNA isolated from *E. turbinata* or bacteria of the present invention. Alternatively, inverse and long-range PCR methods may be used to isolate genes and gene clusters encoding NRPS genes and gene clusters.

Example 5

Combinatorial Biosynthesis

Combinatorial biosynthesis has generally been used in the search for novel molecules with applications as pharmaceuticals or as platforms for combinatorial synthesis. For example, Shen et al. have demonstrated that engineered aromatic or modular PKSs can be used to generate polyketide libraries of different molecular sizes and shapes (Shen et al., 1999). The biosynthetic genes of the present invention, for example genes for ecteinascidin synthesis, could be incorporated into these systems to create derivatives/analogs of ecteinascidins with improved properties such as reduced toxicity/myalgia, greater efficacy etc. (Shen et al., 1999; Xue et al., 1998). Recently, the erythromycin PKS genes have been engineered to effect combinatorial alterations of catalytic activities in the biosynthetic pathway (McDaniel et al., 1999). This has resulted in the successful generation of more than fifty macrolides which would otherwise be impractical to produce through chemical methods. This leads to the creation of libraries of novel "unnatural" natural products exhibiting altered functions (McDaniel et al., 1999). The ecteinascidin NRPS genes could be used in such a system to create analogs of ecteinascidin with improved properties.

The cloned biosynthetic genes presented here have applications in bioprospecting. The cloned PKS genes could be used in PCR, *in situ* hybridizations, etc to isolate novel marine (and terrestrial) NRPS and ecteinascidins which may exhibit novel structures and novel activities (antibacterial, antifungal, anticancer, etc.)

Example 6

The Localization of Bioactive Secondary Metabolites in Sponges, Ascidians and their Associated Symbionts

Methods

Adult *E. turbinata* were shipped by CalBioMarine Technologies Inc., Carlsbad, CA from wild colonies in Florida and maintained in static aquaria at 28°C for up to 2 weeks. Larvae were either released from the colonies in the morning and individually collected by pipette or excised from adult individuals with a scalpel. Some larvae were collected directly in Sweetings Cay, Bahamas by dissecting gravid adults within an hour of collection. Larvae were collected from the adults, either by tank raising colonies or excising gravid adults. Larvae collected by dissection of individuals possessed a range of developmental stages (Figure 5). Larvae were washed several times in sterile seawater prior to all experiments.

Larval chemistry. Three hundred larvae (~300 mg wet weight) were extracted in methanol which was dried *in vacuo* to afford 6.9 mg of extract. This crude extract was dissolved in deuterated methanol (CD_3OD) and analyzed by ^1H NMR spectroscopy. The sample was then dried and dissolved in DMSO at a concentration of 10 mg/mL and tested in a Human Colon Tumor (HCT) bioassay. A portion of the extract was dissolved in methanol (100 µg/mL) and analyzed using ESI mass spectrometry.

PCR of *E. turbinata* bacterial 16S rRNA genes. DNA was extracted from either fresh larvae or larvae preserved in ethanol at 4°C using a DNeasy kit (Qiagen). Portions of the 16S rRNA gene were amplified using "universal" eubacterial primer pairs: 27f-1492r and 1055f-1392r (see Table 1 for all oligonucleotide sequences used) (Amann et al., 1995). PCR reactions were performed with 150-200 ng larval genomic DNA, *Taq* polymerase, 1X Mg buffer, 10 µg BSA, and 250 µM dNTP using a Hybaid thermocycler. A 30 cycle program was used with the following steps: (1) 95°C for 30 seconds for regular primers or 1 minute for GC clamp primers; (2) 50° for 1 minute; (3) 72° for 1 minute; and (4) a final elongations step at 72° for 7 minutes. Prior to

sequencing, all PCR products were cleaned using a QIAQuick™ cleanup kit. Sequencing was performed using ABI PRISM™ Dye Terminator Cycle Sequencing. Sequences were analyzed and aligned using Sequencher™ and by eye.

Table 1

Oligonucleotide	Sequences (5′ → 3′)	Application[a,b]	References
Symbiont specific			
ET1253r	GGG UUG CGA AGC AGC GAU G (SEQ ID NO:2)	F	Herein
Domain Bacteria			
Eub27f	TGA GCC AGG ATC AAA CTC T (SEQ ID NO:3)	P	Amann et al., 1995
Eub338	GCT GCC TCC CGT AGG AGT (SEQ ID NO:4)	F	Fuchs et al., 1998
Eub1055f	ATG GCT GTC GTC AGC T (SEQ ID NO:5)	P	Amann et al., 1995
Eub1392r	CGCCCGCCGCGCCCCGC GCCCGGCCCGCCGCCCC CGCCCCACGGGCGGTGT GTAC (SEQ ID NO:6)	P	Amann et al., 1995
Eub1492r	AAG TCG TAA CAA GGA AGC CGT A (SEQ ID NO:8)	P	Amann et al., 1995
Controls			
Bn1253r	CAT CGC TGC TTC GCA ACC C (SEQ ID NO:9)	F, P	Haygood and Davidson, 1997
Bn240r	TGC TAT TTG ATG AGC CCG CGT T (SEQ ID NO:10)	P	Haygood and Davidson, 1997
Bncontrol	ACG TCA CCG TCC AGC CTC T (SEQ ID NO: 11)	F	Haygood and Davidson, 1997
End 855			
Ep1247r	AAG TAG CGG CCC TTT GTC (SEQ ID NO:12)	P	Schmidt et al., 2000

[a] P = PCR primer, F = FISH probe

[b] Oligonucleotide sequences used for FISH experiments were 5′ biotinylated

Denaturing gradient gel electrophoresis (DGGE) of bacterial 16S rRNA from *E. turbinata*. DGGE is a technique that has been used to separate and identify bacterial sequences from complicated mixtures of microbial populations based on the small subunit rRNA genes (SSU rRNA) (Sakai et al., 1996). Ribosomal RNA is an ideal molecule to use for phylogenetic studies because it is relatively conserved throughout all organisms and contains interspersed variable regions. "Universal" primers are used to amplify the variable regions of the SSU rRNA gene from a mixed population of bacteria. These amplified gene fragments of the same size are then separated electrophoretically according to sequence differences (primarily GC content) on a denaturing gel. The resulting bands can be excised from the gel, reamplified and sequenced.

DGGE experiments were performed on SSU rRNA gene fragments (1055-1392 corresponding to *E. coli* numbering) amplified from *E. turbinata* genomic DNA to study the diversity of associated microbes. PCR products containing a GC clamp were cleaned with a QIAQuick kit. The 8% acrylamide gel gradients were made with 20-70%, 30-70% and 30-60% denaturant (urea) and run in 1X TAE buffer at 60°C for either 5 or 14 hours (200V or 80V, respectively) using a BioRad DCode DGGE system. Bands were stained for 20 minutes with Sybre-gold™ fluorescent stain, documented using a Nucleotech gel documentation system with Nucleovision software and cut out using sterilized razor blades. Each band was incubated in an equivalent volume of 0.5 X TE buffer overnight at 37°C and 5 µL of this solution was used as a template for PCR amplification with 1055f and 1392r primers. PCR products were cleaned, ethanol precipitated and submitted for sequencing.

Microscopy. For TEM, larvae were fixed in Kamovsky's fixative at 4°C. Samples were further fixed in 1% osmium tetroxide in cacodylate buffer for 1 hour, stained with 4% ethanolic uranyl acetate, dehydrated in an ethanol/propylene oxide series, and embedded in Epon resin. Thick sections (1 µm) were obtained on a Reichert Ultramicrotome and stained with toluidene blue. Thin sections (silver-gold by interference) were stained with bismuth

(0.025 µl saturated bismuth per µl water) and 4% ethanolic uranyl acetate and examined on a Zeiss TEM. Light and confocal microscopic images were obtained in the region near the adhesive papillae for orientation.

Immunolocalization experiments. For immunogold experiments, larvae were fixed in paraformaldehyde and then hybridized with the fluorescent EUB 338 probe (see FISH section), dehydrated in an ethanol series, embedded in LR white resin, cured overnight at 50°C, and sectioned to gold thickness by interference. Sections were picked up on parlodion coated nickel grids and incubated on drops of blocking buffer (TBS, pH 7.6, 0.8% BSA, 5% normal goat serum, 0.1% sodium azide) for 30 minutes at room temperature. Grids were transferred to goat anti-fluorescein antibody conjugated to 10 nm colloidal gold diluted 1:25 with gold buffer (TBS, 1% normal goat serum, 0.1% gelatin, 0.2% sodium azide) for 2 hours at room temperature. Grids were washed in buffer (blocking buffer without goat serum) 2 × 15 minutes, TBS 2 × 10 minutes, and distilled water 1 × 5 minutes. Tissues containing immunolinked gold were fixed by incubating sections in 2% glutaraldehyde in 0.1 M Na Cacodylate buffer, pH 7.4 for 10 minutes. Grids were then counterstained with 2% aqueous uranyl acetate for 10 minutes followed by bismuth subnitrate for 10 minutes, and examined by TEM.

Fluorescence *in situ* hybridization (FISH). The probes used for the hybridization experiments were the universal eubacterial probe EUB338 or "symbiont specific" probe Et1253r. The symbiont specific probe was designed based on the sequence obtained from the two DGGE bands (see above). The most variable regions of the 16S gene between 1055 and 1392 were analyzed for site accessibility and similar sequences using CHECK PROBE (RDP) (Amann et al., 1995; Fuchs et al., 1998).

For control experiments, the following probes were used: the bryozoan symbiont specific probe (*Endobugula sertula*) Bn1253r (Haygood and Davidson, 1997); a scrambled probe Bncontrol (Haygood and Davidson, 1997); the sponge symbiont specific probe *Entotheonella palauensis* 1249r (Schmidt et al., 2000);

and no probe controls. The sequences of all probes used are shown in Table 1. Each probe was 5′ labeled with either fluorescein or CY5.

The hybridization protocol was adapted from Haygood and Davidson (Haygood and Davidson, 1997). Samples were fixed in 4% paraformaldehyde/PBS (phosphate buffered saline, pH 7.4), dehydrated through an ethanol series to 70% and stored at 4°C. Approximately 10 larvae were permeabilized with proteinase K, refixed with 4% paraformaldehyde in PBS, and blocked with acetic anhydride. After washing, they were then suspended in pre-hybridization buffer (1 mg total RNA and 100 µg heparin per mL, 10% formamide, 5X SSC, 0.1% Tween 20, 5 mM EDTA and 1X Denhardts' solution) and incubated at 43°C for 1 hour. For the hybridization step, labeled probe was added to a concentration of 2 ng/µl, and the samples were incubated at 55°C for 10 minutes, and then overnight (~15 hours) at 43°C. Samples were then washed according to the following schedule: (1) 2 times for 15 minutes at 43°C with high stringency solution (10% formamide, 5 × SSC, and 1% sodium dodecyl sulfate); (2) 2 times 15 minutes at 37°C with wash 2 (10% formamide, 2 × SSC, and 1% sodium dodecyl sulfate); (3) and once at room temperature for 30 minutes with wash 3 (10% formamide, 2 × SSC, and 0.1% Tween 20). After the final wash, the larvae were suspended and stored in PBS. For analysis, the larvae were mounted in Vectashield™ and examined with a Zeiss 510 confocal laser-scanning microscope. All excitation wavelengths (488, 568, and 647 nm) were used simultaneously to help visualize the tissue surrounding any signal. Images were obtained and manipulated using 3-D for LSM software.

Results

Chemistry. The larval extracts were very active in the HCT bioassay (IC_{50} = 0.65 µg/mL), which suggested the presence of the ecteinascidins. No ecteinascidins could be detected by ^1H NMR analysis of the larval extracts. However, analysis by ESI MS provided a parent ion of 744 which corresponds to Et-743 $[M+H]^+$, and bombardment of that peak produced the fragment 494, which corresponds to the loss of the C-ring subunit (Figure 6).

PCR. The use of universal eubacterial primers allowed amplification of ~1400 bp of bacterial 16s rRNA sequences from genomic DNA isolated from larvae, although the yields were generally low. The amplified products were submitted for sequencing directly, but a clean sequence was not obtained. The ~300 bp PCR products were subjected to DGGE analysis (Figure 7). The two diffuse bands were cut out and sequenced. The consensus rRNA sequence was determined to be, from 5' to 3',

TTGGGTTAAGTCCCGTAACGAGCGCAACCCTTTCCCTTAGTTGCCAGC
GTGTAAAGACGGGGACTCTGAGGGGACTGCCGGTGATAAACCGGAG
GAAGGCGAGGACGACGTCAAGTCATCATGGTCCTTACGAGTAGGGCT
ACACACGTGCTACAATGGTATGTACAAAGGGAGGCAAAATTGTAAAA
TCTAGCAAATCCCCAAAAGCATATCTTAGTCCGGATTGAAGTCTGCA
ACTCGACTTCATGAAGTTGGAATCGCTAGTAATCGCGAATCAGCATG
TCGCGGTGAATACGTTCCCGGGCCTT (SEQ ID NO:13).

Sequence analysis showed that the two bands were almost identical, and different by only four bases/gaps.

Phylogenetic analysis. The ~345 bp sequences obtained from the DGGE bands were submitted to both BLAST (GenBank) and RDP online analysis programs to find the most closely related sequences. The top 8 matches were aligned with the larval sequences and several outgroups using Sequencher™. The related sequences used in the alignment and trees were uncultured gamma proteobacterium clone 8-65, *Bathymodiolus platifrons, Methylobacter psychrophilus, Methylmonas* sp. LW 16, clone L013.6, endosymbiont of the greenhouse whitefly, *Trialeurodes vaporariorum, Bathymodiolus japonicus, Halomonas nitritophilus.* Outgroups were *Vibrio fischeri* and *Streptomyces lavendulae* (Figure 8).

FISH. In the FISH experiments, both the EUB338 and *E. sertula* Bn1253r probes hybridized to specific cells. Little to no signal was detected from the no probe control, *E. palauensis* 1247r and the scrambled Bncontrol probe (Figure 9). The cells marked by signal were quite large (5-10 μm) and

appeared to be ascidian cells. Higher resolution analysis was attempted using thick sections but the cells could not be identified (Figure 10).

Due to the positive result from the *E. sertula* probe, the specificity of the binding was tested using more stringent PCR amplification. PCR products were obtained for all bryozoan samples as expected, and only from the eubacterial primers for the *E. turbinata* samples (Figure 11).

Microscopy. Both light microscopic and TEM analysis of the larvae suggests that there is not a large population of bacteria either within or associated with the host cells, as the FISH experiments appeared to suggest (Figure 12). A few bacteria (1-3) could occasionally be seen by TEM within the larvae, but external to larval cells. The gold-labeled anti-fluorescein antibody experiments failed to localize the cells associated with the original oligo signals.

Discussion

The structural similarity of the ecteinascidins to known microbial metabolites provided compelling circumstantial evidence for a bacterial source of these compounds in the ascidian. By focusing on non-feeding *E. turbinata* larvae, complications such as filter feeding in the adults could be avoided. The secondary metabolites of the larvae were determined in conjunction with genomic studies in an attempt to obtain information about the biosynthetic source of the ecteinascidin compounds.

The chemical analyses showed that Et-743 is present in the larvae in high enough yields that it can be easily detected in crude extracts by ESI mass spectrometry. Although the exact location of the metabolites is not known, it is suspected that at least some are associated with the surface of the larvae due to the immediate deterrent effect against fish upon uptake into the mouth (Bingham, 1986).

Initial PCR experiments using bacterial primers suggested that bacteria were associated with the larvae. FISH experiments using eubacterial probes also suggested the presence of bacterial sequences either inside or associated with ascidian host cells. All the control probes except for the Bn1253r (*Endobugula*

sertula) probe provided negative results. The presence of *E. sertula* or closely related species was tested by PCR amplification using *E. sertula* specific primers. No amplification was detected.

Due to the relatively abundant signal inferred by the eubacterial probe, it was expected that large numbers of bacteria would be detected by TEM analysis. However, extremely few morphologically distinct bacteria could be visualized.

It is possible that there are bacterial symbionts associated with the ascidian larvae in small numbers or with only adults that produce Et-743. There are some invertebrates that contain obligate symbionts as adults but do not transmit them to their offspring. For example, sepiolid squids contain symbiotic bacteria (*Vibrio fischeri*) in their light organs that are responsible for bioluminescent light production, but the symbionts are not vertically transmitted to their eggs. Instead, the young squids selectively obtain the same symbiont from the ambient seawater (Nyholm et al., 2000). Another mode of vertical symbiont transfer is for the adult to brood eggs aposymbiotically and then allow the passage of bacteria onto them as they pass specific organs into the environment (Giere et al., 1991). Some larvae were excised from *E. turbinata* for this study prior to natural release and could therefore be aposymbiotic due to this possible mode of symbiont transfer. However, TEM and FISH experiments conducted with both "natural" larvae and dissected samples provided identical results in all cases.

The ecological and biological functions of Et-743 are still largely unknown, although there is some evidence to suggest that they serve as a feeding deterrent for the larvae (Young and Bingham, 1987). Feeding experiments using adult extracts showed that they were only deterrent when the protein level of the test foods matched the extremely low concentration in the ascidian (Kicklighter, 2000). Other possible functions include prevention of bacterial or viral infection, settlement cues, and/or prevention of encroachment of nearby species. Some workers have also found that the tiger flatworm *Pseudoceros crozeri* feeds exclusively upon adult *E. turbinata* and selectively sequesters some of the ecteinascidins (Wright, 1999).

Example 7

Characterization of Putative Bacterial Symbionts of *Ecteinascidia turbinata*

DNA was isolated from *Ecteinascidia turbinata* (ET) larvae stored in proteinase k (from the Qiagen DNeasy kit) at 4°C. From fresh ETs, numerous larvae were collected for fixing for FISH, RNA isolation, DNA isolation, and DAPI staining. C-sections were performed and numerous larvae in different developmental stages were collected and stored, as well as the adults, in ethanol.

Results

9.5 µg RNA was isolated from various batches of larvae.

DAPI staining showed fluorescence in small round dots of consistent size. A second DAPI staining was performed using ethanol-preserved larvae in order to control for potential auto-fluorescence by using an unstained control. While no autofluorescence was seen in the second DAPI, the dots were not present in the stained samples.

PCR experiments were performed using 130-190 ng of larval DNA and the 1055f/1392r primers without the GC clamp; and, because the 338 probe sequence gave a positive signal in FISH (Example 6), in a separate reaction the 338f/1392r (non-GC clamp) primers were used. PCR conditions were identical to those in Example 6. The PCR products were reamplified using the 1055f/1392r primers, with a GC clamp on the reverse primer, and several reactions produced strong bands.

The products were cleaned with an ethanol precipitation, quantified, and a DGGE with a 30-70% gradient on a 8% gel was performed with both a negative and positive (*E. coli*) control. As in Example 6, samples gave rise to two wide bands on the DGGE gel, in addition to other bands. Seventeen bands were removed from the DGGE gel. DNA from the bands was purified and reamplified with the 1055f/1392r primers. Following clean-up, either 50 ng or 90 ng of DNA DNA was subjected to sequencing. Sequence of the two bands was, 5' to 3',

Band 11c

TATGGCTGTCGTTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGT
AACGAGCGCAACCCTTTCCCTTAGTTGCCAGCGTGTAAAGACGGGGA
CTCTGAGGGGACTGCCGGTGATAAACCGGAGGAAGGCGAGGACGAC
GTCAAGTCATCATGGTCCTTACGAGTAGGGCTACACACGTGCTACAA
TGGTATGTACCAAGGGAGGCAAMATTGTAAAATCTARCAAATCCCC
MAAAGCATATCTTAGTCCGGATTGAAGTCTGCAACTCGACTTCATGA
AGTTGGAATCGCTAGTAATCGCGAATCAGCTATGTCGCGGTGAATAC
GTTCCCGGGCCTTGTACACACCG (SEQ ID NO:14); and

Band 11b

TTAGTTGCAGCGTGTAAGACGGGGACTCTGAGGGGACTGCCGGTGAT
AAACCGGAGGAAGGCGAGGACGACGTCAAGTCATCATGGTCCTTAC
GAGTAGGGCTACACACGTGCTACAATGGTATGTACAAAGGGAGGCA
AAATTGTAAAATCTAGCAAATCCCCAAAAGCATATCTTAGTCCGGAT
TGAAGTCTGCAACTCGACTTCATGAAGTTGGAATCGCTAGTAATCGC
GAATCAGCATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACAC
(SEQ ID NO:15).

The consensus sequence from these bands was exactly as that of Example 6. Using this data, several phylogenetic trees using both distance and parsimony methods were constructed, and in all cases, the DGGE sequences were most closely related to two symbionts of the whitefly as described in Example 6. The sequences were also related to several Halomonas species.

All publications, including patent documents and scientific articles, referred to in this application, including any bibliography, are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication were individually incorporated by reference.

REFERENCES

Amann et al., Microbiol. Rev., 59, 143-169 (1995).

Amann et al., J. Bacteriol., 172, 762-770 (1997).

Arai et al., J. Antibiot., 30, 1015-1019 (1977).

Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons (1998).

Benayahu and Schleyer, Marine Biol., 131, 433-443 (1998).

Benson, Microbial Applications, 6th Ed. Wm. C. Brown Publishers, Dubuque, Iowa (1994).

Bewley et al., Experientia, 52, 716-722 (1996).

Bingham, Am. Zoologist, 26, 127A (1986).

Boyd et al., Drug Development Research, 34, 91-109 (1995).

Corey et al., J. Am. Chem. Soc., 118, 9202-9203 (1996).

Creasey, Drug Disposition in Humans, The Basis of Clinical Pharmacology, Oxford University Press, Oxford (1979).

Cuevas et al., Organic Letters, 2, 2545-2548 (2000).

Daluge et al., Antimicro. Agents Chemother., 41, 1082-1093 (1995).

Davidson and Haygood, Biol. Bull., 196, 273-280 (1999).

Faircloth et al., Eur. J. Cancer, 32A, S5 (1996).

Fontana et al., Tetrahedron, 56, 7305-7308 (2000).

Frincke et al., J. Am. Chem. Soc., 104, 265-269 (1982).

Fuchs et al., Applied and Environmental Microbiology, 64, 4973-4982 (1998).

Garcia-Rocha et al., British J. Cancer, 73, 875-883 (1996).

Giere et al., Marine Ecology - Progress Series, 68, 287-299 (1991).

Gunasekera et al., J. Org. Chem., 55, 4912-4915 (1990).

Hakala, et al., Methods in Cancer Research DeVita and Busch, eds., Academic Press, New York, Vol. XVI (1987).

Harlowe and Lane, Antibodies, a Laboratory Manual, Cold Spring Harbor Press (1988).

Haygood and Davidson, Applied and Environmental Microbiology, 63, 4612-4616 (1997).

He et al., J. Org. Chem., 54, 5822-5824 (1989).

Herberich et al., Tetrahedron Letters, 42, 543-546 (2001).

Hirose, Zool. Sci., 17, 233-240 (2000).

Irschik et al., J. Antibiot., 41, 993-998 (1988).

Kerr et al., J. Nat. Prod., 58, 1618-1621 (1995).

Kicklighter, personal communication (2000).

Konz et al., Chemistry & Biology, 6, R39-R48 (1999).

Krueger et al., Biol. Bull., 190, 195-202 (1996).

Lindel et al., J. Am. Chem. Soc., 119, 8744-8745 (1997).

Lu, Basic Toxicology, Fundamentals, Target Organs, and Risk Assessment, Hemisphere Publishing Corp., Washington (1985).

Marahiel et al., Chem. Rev., 97, 2651-2673 (1997).

Martinez and Corey, Organic Letters, 2, 993-996 (2000).

McDonough, Phys. Ther., 62, 835-839 (1982).

McKee and Ireland, J. Natural Products, 50, 754-756 (1987).

Meyers et al., J. Antibiot., 36, 190-193 (1983).

Mikami et al., J. Biol. Chem., 260, 344-348 (1985).

Moore et al., J. Am. Chem. Soc., 119, 5475-5476 (1997).

Morre et al., J. Am. Chem. Soc., 120, 2490-2491 (1998).

Nyholm et al., Proc. Natl. Acad. Sci. USA, 97, 10231-10235 (2000).

Parameswaran et al., Indian J. of Chem. Section B - Org. Chem. including Medicinal Chem., 37, 1258-1263 (1998).

Petricek et al., FEMS Microbiol. Letts., 61, 109-113 (1989).

Pettit et al., J. Am. Chem. Soc., 104, 6846-6848 (1982).

Pettit et al., J. Natural Products, 63, 793-798 (2000).

Pommier et al., Biochemistry, 35, 13303-13309 (1996).

Pospiech et al., Microbiology, 141, 1793-1803 (1995).

Pospiech et al., Microbiology, 142, 741-746 (1996).

Rinehart et al., J. Org. Chem., 55, 4512-4515 (1990).

Rinehart, Medicinal Research Reviews, 20, 1-27 (2000).

Sakai et al., J. Am. Chem. Soc., 118, 9017-9023 (1996).

Salomon et al., Marine Biol. (2001).

Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989).

Schmidt, Marine sponges and symbionts: Chemical and biological studies, Ph.D. Thesis, University of California, San Diego, 1999.

Schmidt et al., Marine Biol., 136, 969-977 (2000).

Scopes, Protein Purification: Principles and Practice, Springer Verlag, New York (1994); Affinity Chromatography: Principles and Methods, Pharmacia LKB Biotechnology, Uppsala, Sweden (1988).

Shaw and Lacy, J. Bone Joint Surg. (Br.), 55, 197-205 (1973).

Strathmann, Ann. Rev. Ecology and Systematics, 16, 339-361 (1985).

Thompson et al., Acta Zoologica, 64, 199-210 (1983).

Tsoi et al., Chemistry & Biology, 2, 355-362 (1995).

Unson et al., Marine Biol., 119, 1-11 (1994).

Valori et al., Clin. Cancer Res., 4, 1977-1983 (1998).

Watanabe et al., J. Am. Chem. Soc., 120, 6231-6239 (1998).

Wright et al., J. Org. Chem., 55, 4508-4512 (1990).

Wright, personal communication (1999).

Young and Bingham, Marine Biol., 96, 539-544 (1987).

Zewail-Foote and Hurley, J. Medicinal Chem., 42, 2493-2497 (1999).

Zhou et al., Tetrahedron Letters, 41, 2039-2042 (2000).

WHAT IS CLAIMED IS:

1. A composition, comprising a bacteria isolated from *Ecteinascidia turbinata*.

2. The composition of claim 1, wherein said bacteria comprises at least one activity that is involved in the production of an ecteinascidin or a precursor thereof.

3. A composition, comprising: at least one nucleic acid molecule that encodes at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin, wherein said at least one nucleic acid is derived from at least one marine organism.

4. The composition of claim 3, wherein said at least one polypeptide comprises at least one activity of at least one non-ribosomal peptide synthase.

5. The composition of claim 3, wherein said at least one marine organism comprises at least one bacteria.

6. The composition of claim 3, wherein said at least one marine organism comprises at least one invertebrate.

7. The composition of claim 6, wherein said at least one invertebrate comprises at least one *Ecteinascidia*.

8. The composition of claim 7, wherein said at least one *Ecteinascidia* is *Ecteinascidia turbinata*.

9. The composition of claim 3, wherein said at least one nucleic acid molecule further comprises at least one expression control sequence.

10. The composition of claim 3, wherein said nucleic acid molecule is in a vector.

11. The composition of claim 10, wherein said vector is within a cell.

12. The composition of claim 3, wherein said at least one nucleic acid molecule is within a cell.

13. A composition, comprising a library of nucleic acid molecules of claim 3.

14. A composition, comprising: at least one polypeptide that catalyzes at least one step in the synthesis of at least one ecteinascidin, wherein said at least one polypeptide is derived from at least one marine organism.

15. The composition of claim 14, wherein said at least one ecteinascidin comprises Et-743.

16. The composition of claim 14, wherein said at least one marine organism comprises at least one bacteria.

17. The composition of claim 14, wherein said at least one marine organism comprises at least one invertebrate.

18. The composition of claim 17, wherein said at least one invertebrate comprises at least one *Ecteinascidia*.

19. The composition of claim 18, wherein said at least one *Ecteinascidia* is *Ecteinascidia turbinata*.

20. The composition of claim 14, wherein said at least one polypeptide is within a cell.

21. A composition, comprising a library of polypeptides of claim 14.

22. A composition comprising an antibody or an active fragment thereof that binds with or specifically binds with a bacteria of claim 1.

23. A composition comprising an antibody or an active fragment thereof that binds with or specifically binds with a polypeptide of claim 14.

24. A method of making a composition, comprising: providing at least one composition of claim 1, and synthesizing at least one ecteinascidin.

25. A composition made by the method of claim 24.

26. The composition of claim 25, wherein said composition does not comprise a known ecteinascidin.

27. The composition of claim 24, comprising at least one pharmaceutically acceptable carrier.

28. The composition of claim 24, wherein said composition is a pharmaceutical composition.

29. A method of making a composition, comprising: providing at least one composition of claim 3, and synthesizing at least one ecteinascidin.

30. A composition made by the method of claim 29.

31. The composition of claim 30, wherein said composition does not comprise a known bryostatin.

32. The composition of claim 30, comprising at least one pharmaceutically acceptable carrier.

33. The composition of claim 30, wherein said composition is a pharmaceutical composition.

34. A method of making a composition, comprising: providing at least one composition of claim 14, and synthesizing at least one ecteinascidin.

35. A composition made by the method of claim 34.

36. The composition of claim 35, wherein said composition does not comprise a known ecteinascidin.

37. The composition of claim 35, comprising at least one pharmaceutically acceptable carrier.

38. The composition of claim 35, wherein said composition is a pharmaceutical composition.

39. A method for identifying at least one nucleic acid molecule encoding at least one polypeptide that catalyzes at least one step in the synthesis of an ecteinascidin , comprising: contacting a nucleic acid molecule of claim 1 with a sample, and identifying nucleic acid molecules in said sample that hybridize with said nucleic acid molecule of claim 1.

40. The method of claim 39, wherein said sample is derived at least in part from an environmental sample.

41. The method of claim 40, wherein said environmental sample is derived at least in part from a marine environment.

42. A nucleic acid molecule identified by the method of claim 39.

43. The nucleic acid molecule of claim 42, comprising an expression control sequence.

44. The nucleic acid molecule of claim 42 in a vector.

45. The nucleic acid molecule of claim 42 in a cell.

46. A composition comprising a library of nucleic acid molecules of claim 42.

47. A method for identifying a bioactive compound, comprising: contacting a composition of claim 1 with an *in vitro*, *ex vivo* or *in vivo* detection system and determining the bioactivity of said compound.

48. A bioactive compound identified by the method of claim 47.

49. The bioactive compound of claim 48 in a pharmaceutically acceptable carrier.

50. The bioactive compound of claim 48, wherein said bioactive compound is a pharmaceutical compound.

51. A method for identifying a bioactive compound, comprising: contacting a composition of claim 3 with an *in vitro*, *ex vivo* or *in vivo* detection system and determining the bioactivity of said compound.

52. A bioactive compound identified by the method of claim 51.

53. The bioactive compound of claim 52 in a pharmaceutically acceptable carrier.

54. The bioactive compound of claim 52, wherein said bioactive compound is a pharmaceutical compound.

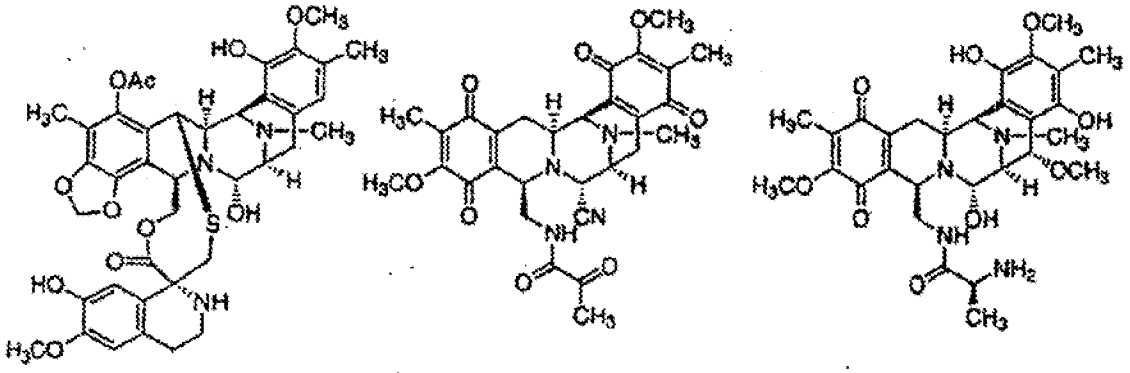
55. A method for identifying a bioactive compound, comprising: contacting a composition of claim 14 with an *in vitro*, *ex vivo* or *in vivo* detection system and determining the bioactivity of said compound.

56. A bioactive compound identified by the method of claim 55.

57. The bioactive compound of claim 56 in a pharmaceutically acceptable carrier.

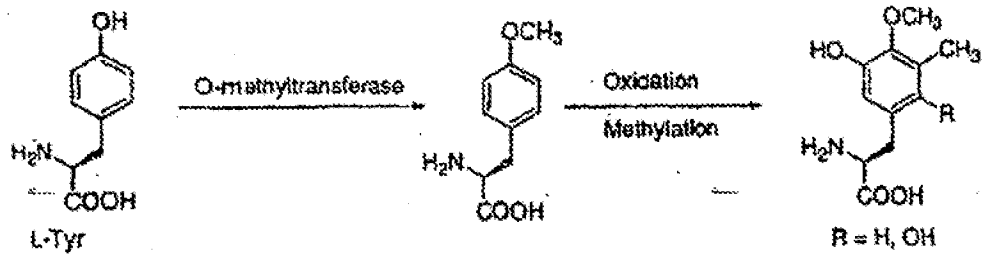
58. The bioactive compound of claim 56, wherein said bioactive compound is a pharmaceutical compound.

Figure 1



Ecteinascidin-743 Saframycin A Saframycin MX1

Figure 2


OH
H2N
COOH
L-Tyr
O-methyltransferase
OCH3
H2N
COOH
Oxidation
Methylation
OCH3
HO
CH3
R
H2N
COOH
R = H, OH

Scheme 1

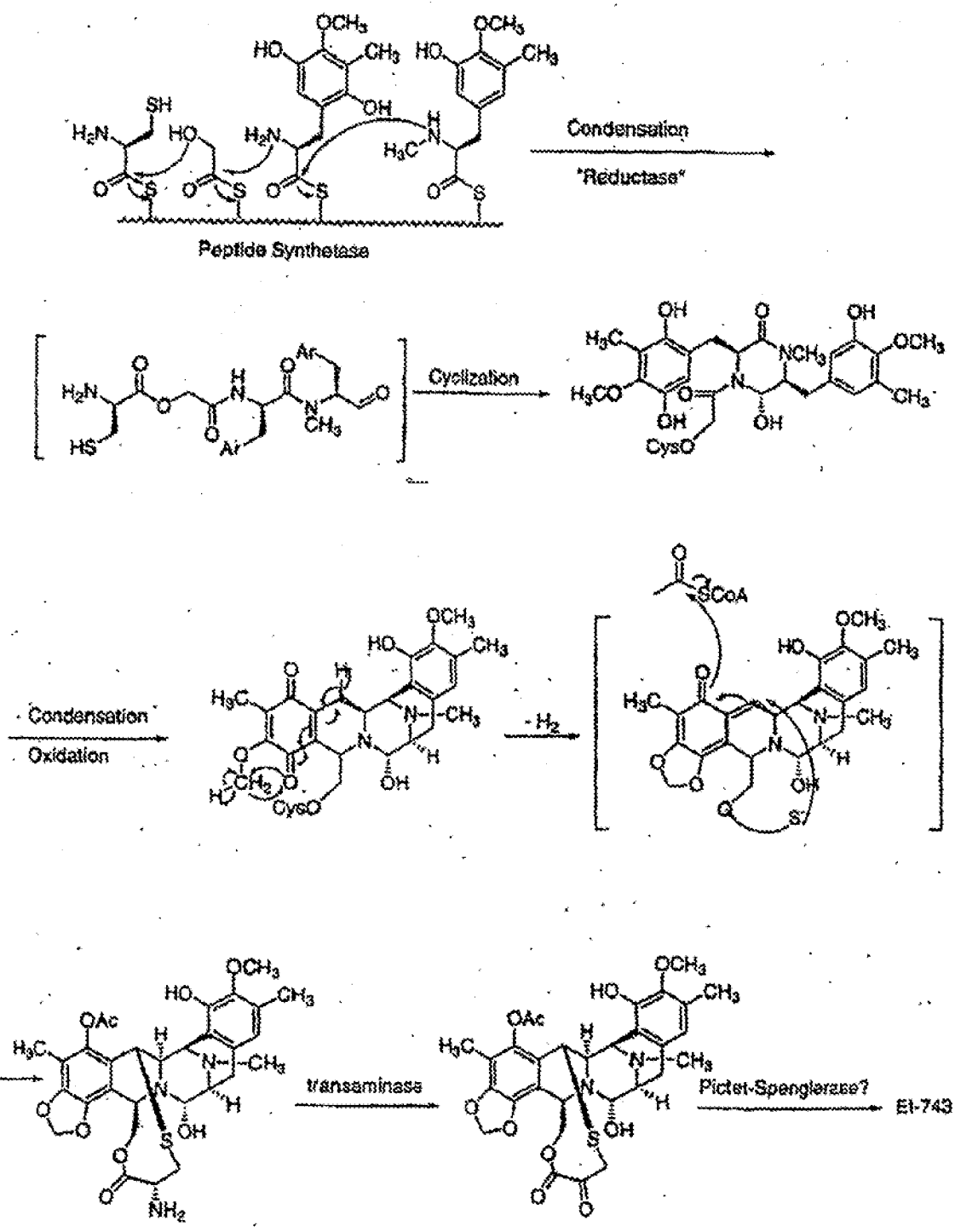
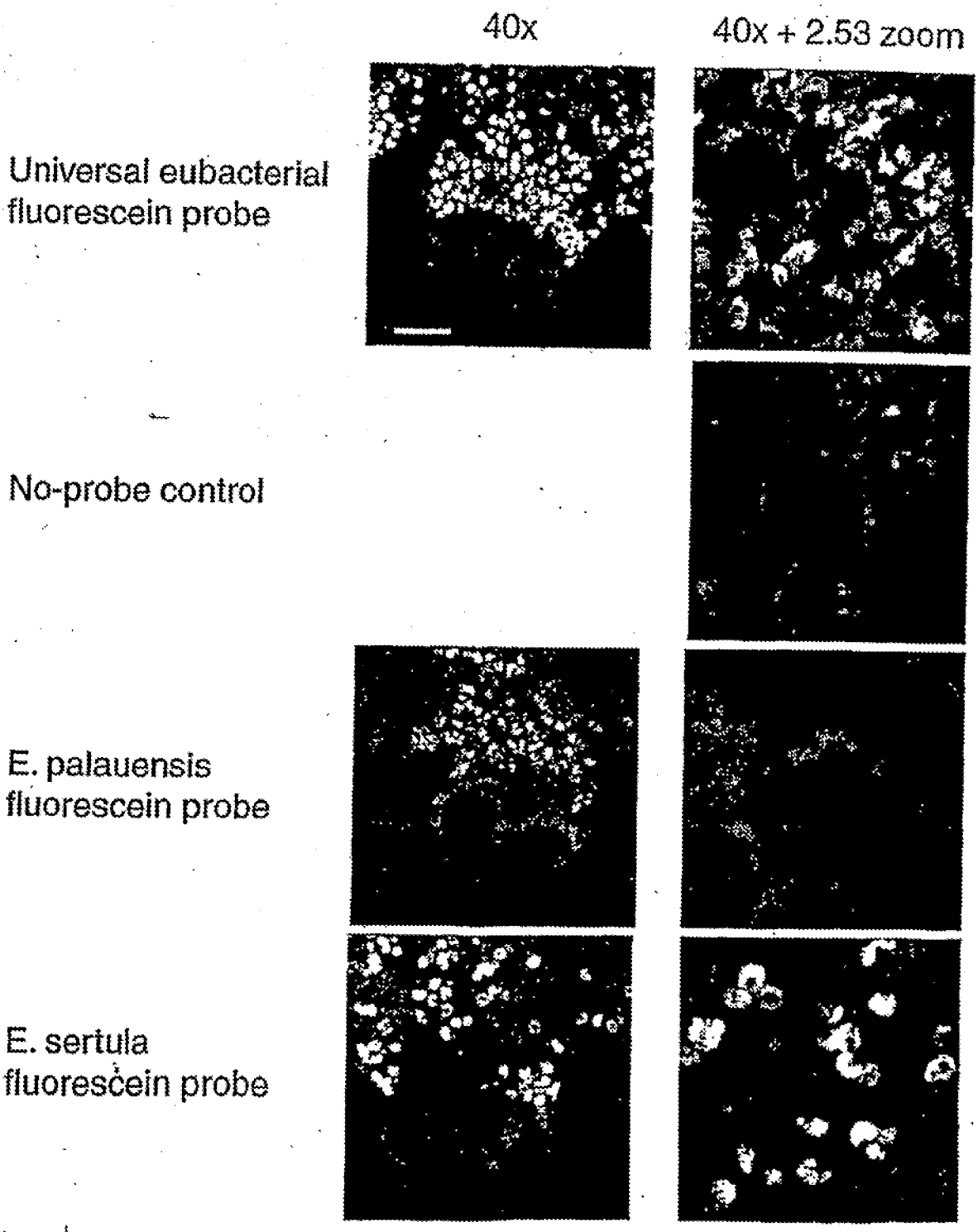


Figure 3

Figure 4



E. turbinata larvae hybridized with bacterial rRNA probes

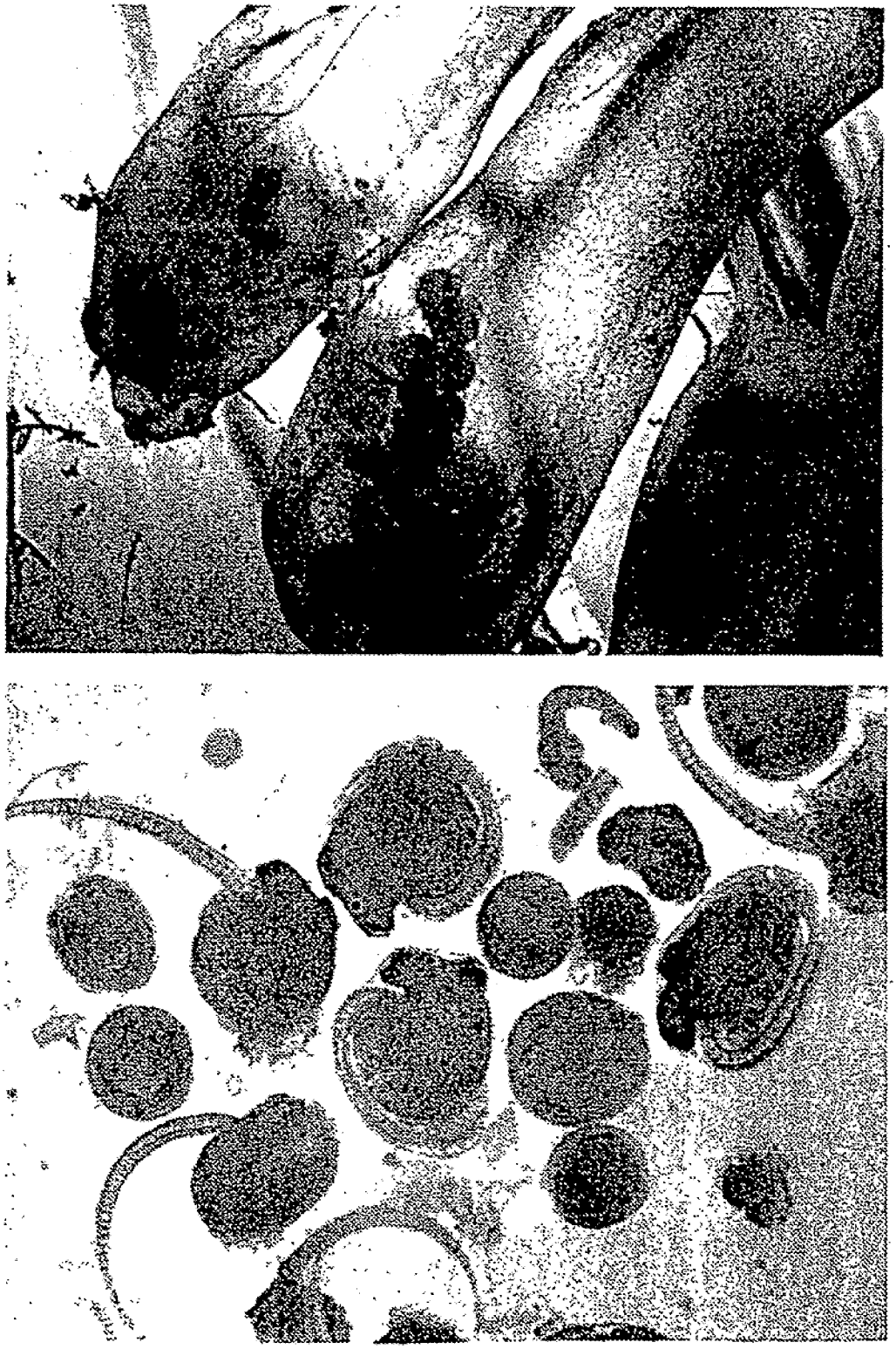


Figure 5

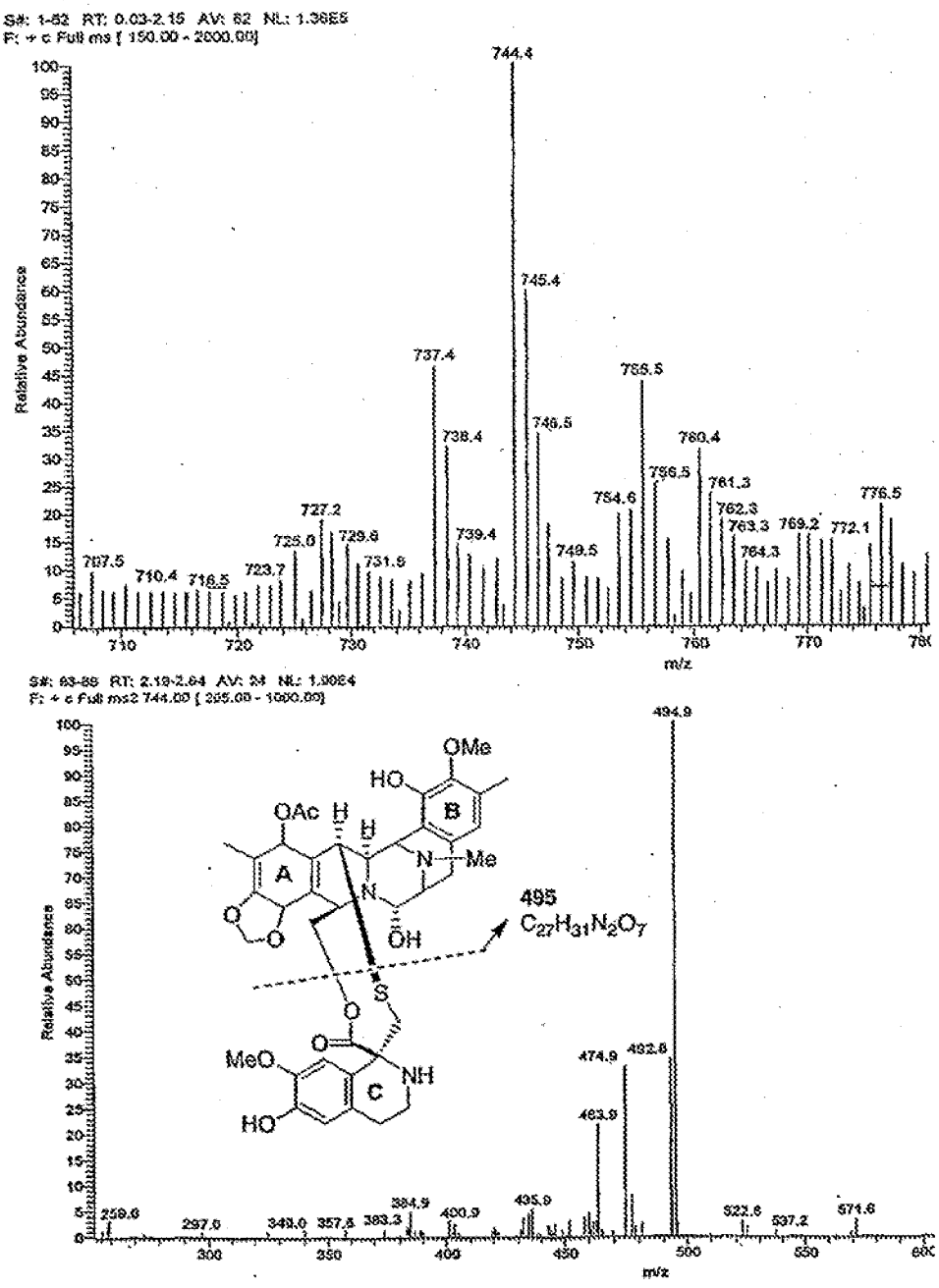

S#: 1-52 RT: 0.03-2.15 AV: 52 NL: 1.36E6
F: + c Full ms [150.00 - 2000.00]
Relative Abundance
m/z
744.4
745.4
737.4
755.5
746.5
738.4
760.4
756.5
761.3
754.6
762.3
727.2
776.5
725.6
729.6
763.3
769.2
772.1
739.4
731.9
749.5
764.3
707.5
710.4
716.5
723.7
S#: 63-69 RT: 2.19-2.64 AV: 34 NL: 1.90E4
F: + c Full ms2 744.00 [255.00 - 1060.00]
494.9
OMe
HO
OAc
H
H
B
A
N
Me
N
O
O
OH
495
C27H31N2O7
S
O
O
MeO
NH
C
HO
474.9
492.8
463.9
259.6
297.5
343.0
357.5
383.3
394.9
400.9
435.9
522.6
537.2
571.6

Figure 6

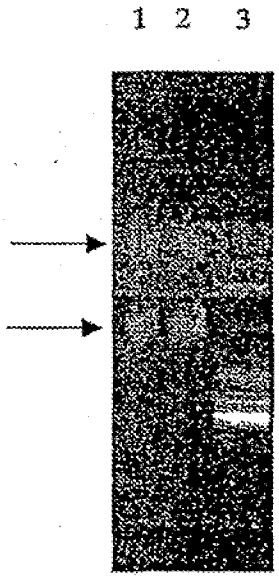


Figure 7

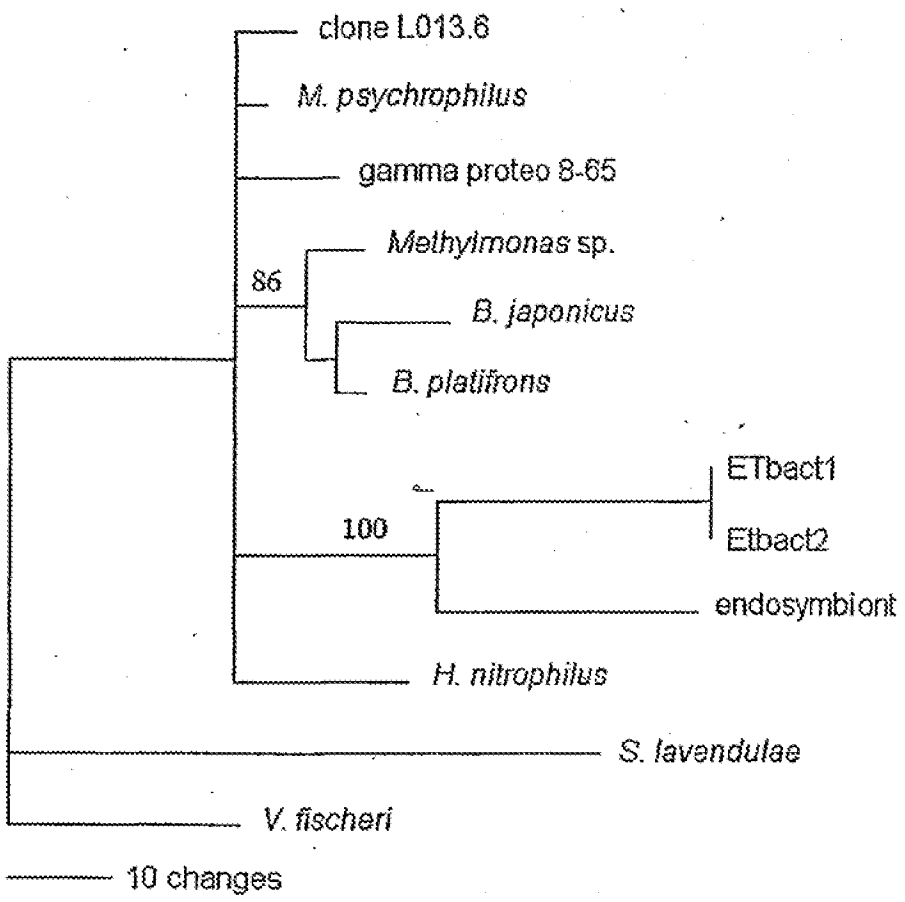
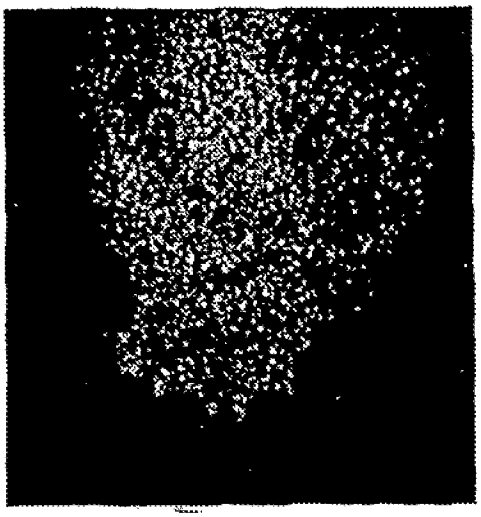


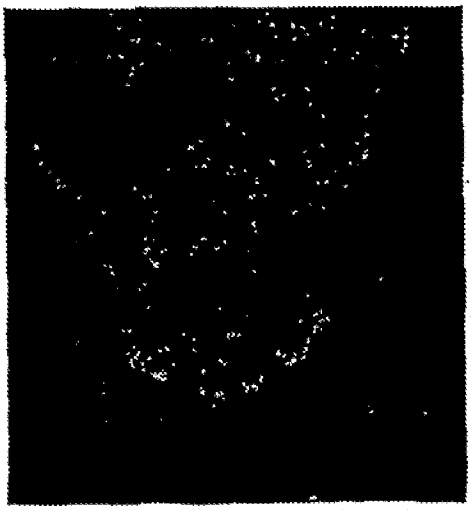
Figure 8

A.



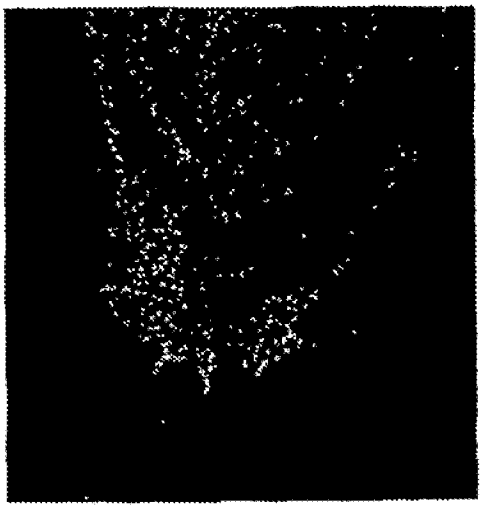
Universal eubacterial fluorescein probe

B.



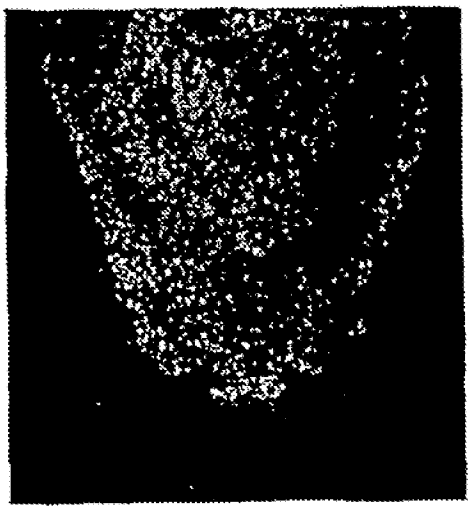
No-probe control

C.



E. palauensis fluorescein probe

D.



E. sertula fluorescein probe

Figure 9

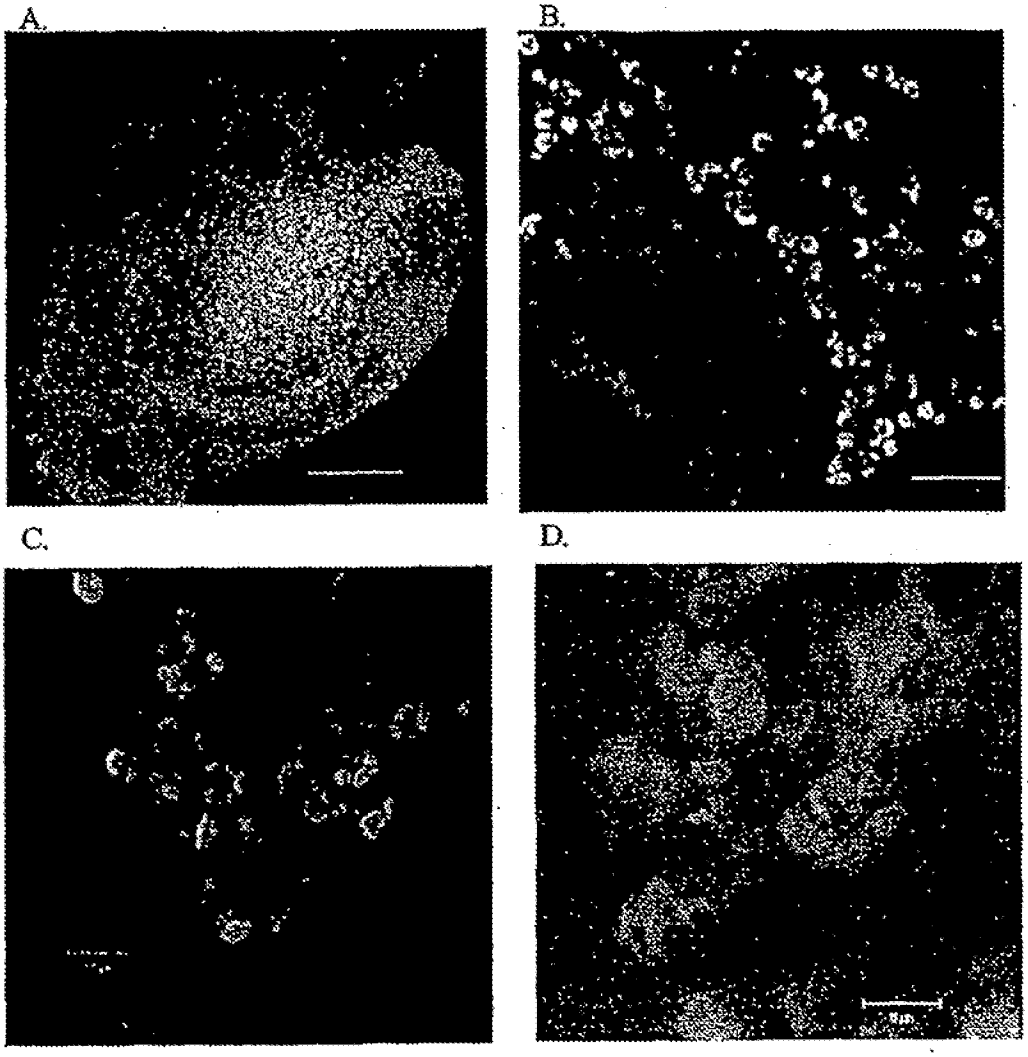


Figure 10

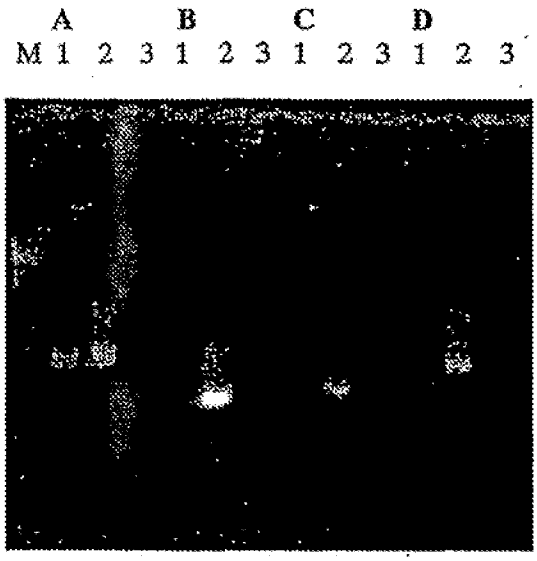


Figure 11

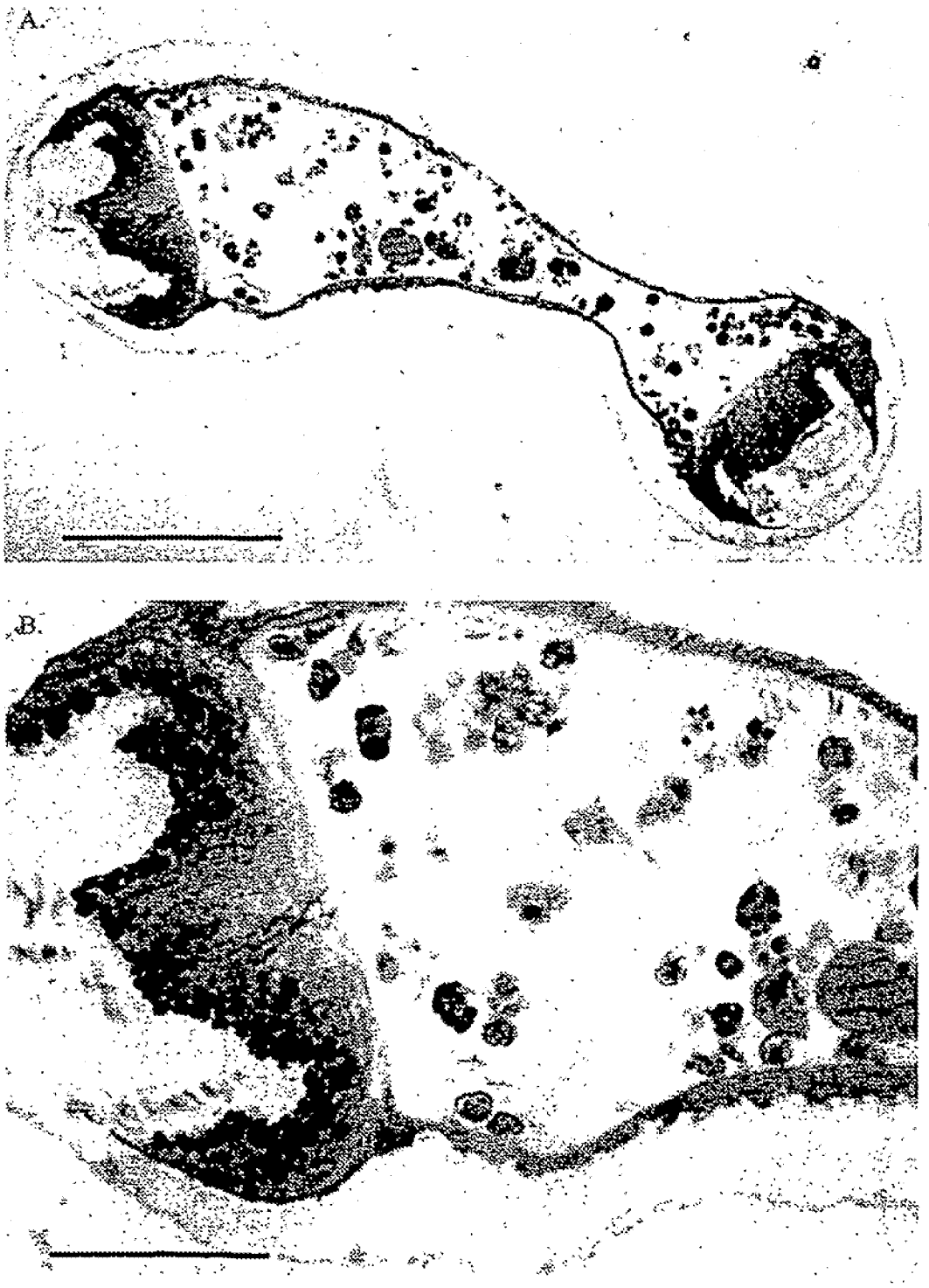


Figure 12

INTERNATIONAL SEARCH REPORT

Int onal Application No
PCT/US 02/03703

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12R1/00 C07D241/36 C12N15/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12R C07D C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JEEDIGUNTA SHANTI ET AL: "Diketopiperazines as advanced intermediates in the biosynthesis of ecteinascidins." TETRAHEDRON, vol. 56, no. 21, 19 May 2000 (2000-05-19), pages 3303-3307, XP002204116 ISSN: 0040-4020 see the whole document, esp. abstract --/--	1-24,29, 34, 39-41, 47,51,55

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 July 2002	30/07/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Grosskopf, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	HAYGOOD M G ET AL: "MICROBIAL SYMBIONTS OF MARINE INVERTEBRATES: OPPORTUNITIES FOR MICROBIAL BIOTECHNOLOGY" JOURNAL OF MOLECULAR MICROBIOLOGY AND BIOTECHNOLOGY, HORIZON SCIENTIFIC PRESS, WYMONDHAM,, GB, vol. 1, no. 1, August 1999 (1999-08), pages 33-43, XP000943299 ISSN: 1464-1801 see the whole document	1-24,29, 34, 39-41, 47,51,55
Y	POMPONI S A: "The bioprocess-technological potential of the sea" JOURNAL OF BIOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 70, no. 1-3, 30 April 1999 (1999-04-30), pages 5-13, XP004173379 ISSN: 0168-1656 see the whole document	1-24,29, 34, 39-41, 47,51,55
A	MUNRO M H G ET AL: "The discovery and development of marine compounds with pharmaceutical potential" JOURNAL OF BIOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 70, no. 1-3, 30 April 1999 (1999-04-30), pages 15-25, XP004173380 ISSN: 0168-1656	
A	FAULKNER D JOHN ET AL: "Symbiotic bacteria in sponges: Sources of bioactive substances." DRUGS FROM THE SEA, 2000, pages 107-119, XP001088056 2000 S. Karger Publishers Inc.;S. Karger AG 26 West Avon Road, Farmington, CT, 06085, USA; CH-4009, Basel, Switzerland ISBN: 3-8055-7098-8	

BNSDOCID: <WO_____03064843A1_I_>

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 25-28, 30-33, 35-38, 42-46, 48-50, 52-54, 56-58

A meaningful search for all of the claims over their whole breadth was not possible due to the absence of any technical features which might be suitable to characterise the essential parts of the compositions and due to the lack of suitable experimental data in the application. Moreover, claims which relate to compositions, nucleic acids, bioactive substances etc. which are merely characterised by the fact that they are "made" by using an undefined composition, cannot be searched at all.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

ernational application No.
PCT/US 02/03703

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 25-28, 30-33, 35-38, 42-46, 48-50, 52-54, 56-58
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.